

Manuscript version: Published Version

The version presented in WRAP is the published version (Version of Record).

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/90346>

How to cite:

The repository item page linked to above, will contain details on accessing citation guidance from the publisher.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

PROOF OF KOMLÓS'S CONJECTURE ON HAMILTONIAN SUBSETS

JAEHOON KIM, HONG LIU, MARYAM SHARIFZADEH AND KATHERINE STADEN

ABSTRACT. Komlós conjectured in 1981 that among all graphs with minimum degree at least d , the complete graph K_{d+1} minimises the number of Hamiltonian subsets, where a subset of vertices is Hamiltonian if it contains a spanning cycle. We prove this conjecture when d is sufficiently large. In fact we prove a stronger result: for large d , any graph G with average degree at least d contains almost twice as many Hamiltonian subsets as K_{d+1} , unless G is isomorphic to K_{d+1} or a certain other graph which we specify.

1. INTRODUCTION

A cycle in a graph G is *Hamiltonian* if it spans the whole vertex set of G . The Hamiltonian cycle problem, i.e. deciding whether a given graph contains a Hamiltonian cycle, is one of Karp's original NP-complete problems [12]. It has been studied in various directions by numerous researchers over the last 60 years. As a computationally hard problem, the *extremal problem* of finding best-possible sufficient conditions that guarantee the existence of a Hamiltonian cycle is of great interest. The classical theorem of Dirac [9], giving a minimum degree condition, was one of the first such results. There have been some exciting recent developments in this area, for example in finding optimal packings of Hamiltonian cycles [14, 16, 23], decompositions of graphs into edge-disjoint Hamiltonian cycles [7, 22], and finding Hamiltonian cycles in hypergraphs (see the survey paper [24] for more references).

The *enumerative problem* of counting the number of cycles in a given class of graphs is one of the oldest problems in graph theory. In 1897, Ahrens [1] proved that, for any graph G on n vertices, if we denote by $\nu(G)$ the number of cycles in G , then

$$e(G) - n + 1 \leq \nu(G) \leq 2^{e(G)-n+1} - 1.$$

Volkman [36] proved that, when G has minimum degree δ , then $\nu(G) \geq \delta(\delta + 1)$. Various authors have investigated the problem of maximising or minimising the number of cycles in other particular classes of graphs, e.g. [3, 5].

Another direction of research is to count the number of Hamiltonian cycles in a graph G . Suppose that $\delta(G) \geq |G|/2$ (and hence G contains at least one Hamiltonian cycle, by Dirac's theorem). Then there is a formula due to Cuckler and Kahn [8] which asymptotically determines the logarithm of the number of Hamiltonian cycles.

In this paper we consider a question which is both extremal and enumerative, namely minimising in a graph G the number $c(G)$ of *distinguishable* cycles, where two subgraphs G_1, G_2 of G are distinguishable if $V(G_1) \neq V(G_2)$. An equivalent formulation, and the one which we shall mainly use, is that $c(G)$ is the number of *Hamiltonian subsets* of G . Here, $A \subseteq V(G)$ is *Hamiltonian* if $G[A]$ contains a Hamiltonian cycle, i.e. G contains a cycle whose vertex set is A . Thus, in contrast to the Hamiltonian cycle problem where one is interested in the collection of cycles which span a given set $V(G)$, we investigate the collection of sets which are spanned by a cycle.

As Dirac's theorem relates the existence of a Hamiltonian cycle to minimum degree, a very natural question is to ask how $c(G)$ relates to minimum degree; that is, minimising $c(G)$ given $\delta(G) \geq d$ for some integer d . Here, the number of vertices n of G is not fixed. Now, every subset of size at least three is a candidate for a Hamiltonian subset, and there are almost 2^n of these. Thus, as n increases, the number of candidates increases, and one expects that $c(G)$ will also increase. Therefore it is natural to conjecture that, given the stipulation $\delta(G) \geq d$, to minimise $c(G)$ one should minimise

Date: October 21, 2018.

J.K. was supported by ERC grant 306349; H.L. was supported by EPSRC grant EP/K012045/1, ERC grant 306493 and the Leverhulme Trust Early Career Fellowship ECF-2016-523; M.Sh. and K.S. were supported by ERC grant 306493.

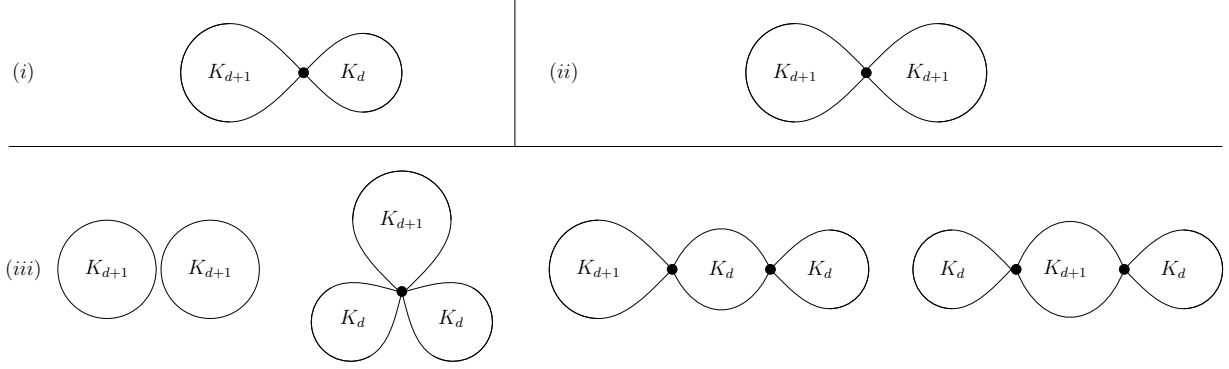


FIGURE 1. (i) the graph $K_{d+1} * K_d$; (ii) a graph G with $d(G) \approx d + \frac{1}{2}$ and $c(G) = (2 - o(1))2^{d+1}$; (iii) four graphs G with $d(G) = d$ and $c(G) = (2 - o(1))2^{d+1}$.

n . Clearly the unique graph G of minimum order with $\delta(G) \geq d$ is the complete graph K_{d+1} . This is the substance of a 1981 conjecture of Komlós (see [33, 34, 35]):

Conjecture 1.1 (see [33]). *For all integers $d > 0$ and all graphs G with $\delta(G) \geq d$, we have*

$$c(G) \geq c(K_{d+1}).$$

Since every subset of size at least three is Hamiltonian in a complete graph, we have

$$c(K_{d+1}) = 2^{d+1} - \binom{d+1}{2} - d - 2. \quad (1.1)$$

In this paper, we prove Komlós's conjecture for all sufficiently large d . In fact, we prove a stronger result which replaces minimum degree with average degree (where the average degree $d(G)$ of a graph G satisfies $d(G)|G| = 2e(G)$), and shows that the extremal graph is stable in a rather precise sense. The current best bound on $c(G)$ is due to Tuza [33].

Theorem 1.2 ([33]). *For all $d \geq 3$ and every graph G with $d(G) \geq d$, we have that $c(G) \geq 2^{d/2}$.*

1.1. New results. The purpose of this paper is to prove Komlós's conjecture for all large d . As mentioned above, we prove a stronger result. To state it, we need a definition. Given any positive integer d , let $K_{d+1} * K_d$ denote the graph obtained by taking vertex-disjoint copies of K_{d+1} and K_d , and identifying them at a single vertex (see Figure 1(i)). Note that $K_{d+1} * K_d$ has average degree exactly d . Moreover,

$$c(K_{d+1} * K_d) = c(K_{d+1}) + c(K_d) = \frac{3}{2} \cdot 2^{d+1} - d^2 - 2d - 3.$$

Our main result is the following.

Theorem 1.3. *For all $0 < \alpha \leq 1$ there exists $d_0 > 0$ such that for all integers $d \geq d_0$, the following holds. Let G be a graph with average degree at least d which is not isomorphic to K_{d+1} or $K_{d+1} * K_d$. Then $c(G) \geq (2 - \alpha)2^{d+1}$.*

This theorem is best possible in the sense that the constant 2 cannot be improved. Indeed, there are many graphs showing that 2 is best possible. Figure 1(iii) shows four graphs G with average degree at least d such that $c(G) = (2 - o(1))2^{d+1}$; and any $(d+2)$ -vertex graph G with average degree at least d also satisfies $c(G) = (2 - o(1))2^{d+1}$. Also the statement is not necessarily true when d is not an integer. The graph G in Figure 1(ii) has average degree slightly less than $d + \frac{1}{2}$ but again satisfies $c(G) = (2 - o(1))2^{d+1} < (2 - o(1))2^{d(G)+1}$; and $K_{d+2} - e$ has average degree slightly less than $d + 1$ but again satisfies $c(G) = (2 - o(1))2^{d+1} < (2 - o(1))2^{d(K_{d+2}-e)+1}$.

A further remark is that Theorem 1.3 is *not true* for $d = 2$. Indeed, every graph with average degree at least two contains a cycle, while $c(C_n) = 1$ for all n . So there are infinitely many graphs which minimise $c(G)$. (Note that this does not contradict Conjecture 1.1.) Although we are not aware of a similar occurrence for any $d \geq 3$, it would be interesting to determine the minimum d_0 one can take in the statement of Theorem 1.3.

Our techniques also imply for large d the following conjecture of Tuza [35], a bipartite analogue of Conjecture 1.1. Since the proof is very similar to the proof of Theorem 1.3, we sketch the details in Section 7.

Theorem 1.4. *There exists $d_0 > 0$ such that, for all integers $d > d_0$ and all bipartite graphs G with minimum degree at least d , we have $c(G) \geq c(K_{d,d})$.*

The graphs G in Conjecture 1.1 could have any number $n \geq d + 1$ of vertices, and consequently G could have any given density. This makes the problem difficult to attack as many available tools depend on the density of the graph. The notions of pseudorandomness and expansion have both played a major role in recent advances in the Hamiltonian cycle problem. In particular, the Regularity-Blow-up method of Komlós, Sárközy and Szemerédi [17], which involves partitioning G into pseudorandom subgraphs, has been the key tool in the solution of many important dense graph problems. For problems involving sparser graphs, the concept of ‘sparse expansion’, introduced by Komlós and Szemerédi [20, 21] has proved very effective (see also [4, 27, 28] for some recent results in which such expanders play a role). A novel aspect of our proof is to combine both approaches; depending, roughly speaking, on whether our graph G is dense or sparse.

The main ingredient in the sparse case is the following general theorem about expander graphs, which we state here as it may be of independent interest. Roughly speaking, it states that an expander graph G which is almost d -regular and not too sparse contains a set Z of order $200d$ such that for every half-sized subset U of Z , there is a cycle in G whose intersection with Z contains almost every vertex in U and no vertices in $Z \setminus U$.

Theorem 1.5. *For given $0 < \varepsilon_1 \leq 1$ and $L \geq 1$, there exist d_0 and K_0 such that the following holds for any $d \geq d_0$, $K \geq K_0$ and $n \in \mathbb{N}$ with $\log^{100} n \leq d \leq n/K$. If H is an n -vertex $(\varepsilon_1, d/30)$ -expander with $d/10 \leq \delta(H) \leq \Delta(H) \leq Ld$, then $V(H)$ contains a set Z of size $200d$ such that, for every subset $U \subseteq Z$ of size $100d$, there exists a cycle C_U with $V(C_U) \cap Z \in \binom{U}{\geq 98d}$.*

We defer the definition of an expander graph to Section 5 and the other notation in the statement to Section 3.

1.2. Related research. Considering a different local condition to that of minimum degree, Tuza [33] proved that, if every edge of a graph G lies in t copies of K_3 , then $c(G) \geq c(K_{t+2})$. In [33, 34, 35], he also considered the problem of, for a given graph F and class \mathcal{G} of graphs, minimising the number of distinguishable subdivisions of F in a graph $G \in \mathcal{G}$. Note that this is a generalisation since a cycle is a subdivision of K_3 . The case when \mathcal{G} is the set of graphs G in which every edge lies in t copies of F is addressed in [26]. Considering a global condition, Knor [15] provided estimates on the number of Hamiltonian subsets in k -connected graphs G for various values of k , in terms of the number of vertices of G .

Given our resolution of Komlós’s conjecture for large d and the fact that the graphs which minimise $c(G)$ are small, it makes sense to ask: what is the minimum value of $c(G)$ over all n -vertex graphs G with minimum (or average) degree d ? A related question of Perkins [30] is minimising the normalised parameter $c(G)/|G|$, which penalises small graphs.

Question 1.6. *What is $f(d) := \liminf_{n \rightarrow \infty} \left\{ \frac{c(G)}{n} : \delta(G) \geq d \text{ and } |G| = n \right\}$?*

For $d = 2$ we have that $c(G) \geq 1$ with equality if and only if $G = C_n$ is a cycle. Thus $f(2) = 0$. For general integers d , a simple lower bound for $f(d)$ can be obtained as follows. Given any graph G with $\delta(G) \geq d$ on n vertices, let $F \subseteq G$ be a forest such that each component spans a component of G . Then, for any $e \in E(G) \setminus E(F)$, the graph $F \cup \{e\}$ contains precisely one cycle, which necessarily contains e . So $c(G)/|G| \geq (e(G) - e(F))/|G| \geq d/2 - 1$. For an upper bound, when $(d+1)|n|$, the union G of $\frac{n}{d+1}$ vertex-disjoint $(d+1)$ -cliques has $c(G) = \frac{n}{d+1} \cdot c(K_{d+1})$. Thus

$$\frac{d}{2} - 1 \leq f(d) \leq \frac{c(K_{d+1})}{d+1}.$$

The upper bound is not tight. This can be seen by taking the vertex-disjoint union of $(d+1)$ -cliques and a cycle C of appropriate length, and adding an edge between each clique and C to ensure $\delta(G) = d$.

What about the problem of *maximising* the number of (distinguishable) cycles? For a fixed number of vertices n , clearly K_n contains the most. If instead we fix maximum degree Δ , then $\nu(G)$ and $c(G)$ both increase as the number of vertices in G increases. To avoid such trivialities, Tuza asked a more restrictive question: what is the maximum number $\nu^{\text{ind}}(G)$ of *induced* cycles in an n -vertex graph G ? Note that any two induced cycles in a graph are distinguishable. This conjecture was verified in a strong sense by Morrison and Scott [29], who showed that there is a unique graph attaining the maximum for every sufficiently large n (which depends on the value of $n \bmod 3$).

Another way of avoiding such trivialities is to ask for the maximum number of cycles in an n -vertex graph G in a restrictive class of graphs. Recently, Arman, Gunderson and Tsaturian [2] showed that $\nu(K_{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor})$ is maximum among all n -vertex triangle-free graphs, for $n \geq 141$.

1.3. Organisation of the paper. The remainder of the paper is organised as follows. In Section 2, we sketch the proof of Theorem 1.3. Our notation and some tools needed for the proof are listed in Section 3. Sections 4 and 5 form the bulk of proof and deal with the ‘dense’ and ‘sparse’ cases respectively (see Section 2 for explanations of these terms). In Section 6, we combine our auxiliary results to complete the proof of Theorem 1.3. Finally, in Section 7, we sketch how to prove Theorem 1.4.

2. SKETCH OF THE PROOF OF THEOREM 1.3

Suppose that G is a counterexample to Theorem 1.3 whose order n is minimal for a fixed large number d . So $d(G) \geq d$ and $G \notin \{K_{d+1}, K_d * K_{d+1}\}$. One can show that the minimality of G implies that it has minimum degree at least $d/2$, and any subgraph has average degree at most d . From now on we distinguish two cases according to whether n/d is bounded (the *dense case*) or unbounded (the *sparse case*). Each case uses different techniques and methods and are essentially separate.

2.1. The dense case: n/d is bounded (Section 4). If $n < 1.19d$ then a simple probabilistic argument proves Theorem 1.3 (see Lemma 4.1). Indeed, a Chernoff bound implies that almost every subset of G is Hamiltonian, via the classical degree sequence theorem of Pósa [31]. Note that 1.19 is somewhat arbitrary here.

Thus we may assume that $1.19d \leq n \leq Kd$ for some (large) constant K which does not depend on d . Since we are assuming that d is large, G is a large dense graph, and we can employ the celebrated Regularity-Blow-up method pioneered by Komlós, Sárközy and Szemerédi [17].

The regularity lemma of Szemerédi [32] implies that G has a partition into equally-sized clusters V_1, \dots, V_r of vertices such that almost all ordered pairs of clusters induce a pseudorandom subgraph of G ; and an exceptional set V_0 of size $o(n)$. The ‘reduced graph’ R of G has vertices $1, \dots, r$, where ij with $1 \leq i < j \leq r$ is an edge if $G[V_i, V_j]$ is pseudorandom and dense. The idea here is to show that R contains some special structure Q with the property that

- (†) there are at least $(2 - o(1))2^{d+1}$ choices of $\{V'_i : i \in V(Q)\}$ where $V'_i \subseteq V_i$ for all i , such that $\bigcup_{i \in V(Q)} V'_i$ is a Hamiltonian subset.

Clearly this will prove Theorem 1.3 in this case. Let $m := |V_1| = \dots = |V_r|$ and let

$$d' := (1 - o(1)) \frac{dr}{n} = (1 - o(1)) \frac{d}{m}. \quad (2.1)$$

There are three possible structures Q we can guarantee in R :

- (1) two vertex-disjoint cycles whose orders sum to at least $1.8d'$;
- (2) a path of length $(1 + \frac{1}{100})d'$;
- (3) a large ‘sun’ with large ‘corona’ (the definitions of which we defer to Section 4.2).

The reason we can find such a structure Q is as follows. It is well-known that the reduced graph R inherits many of the properties of G . In particular, the average degree and minimum degree of R are closely related to those of G . Thus R has average degree at least d' and minimum degree at least $d'/2$. Moreover, $r \geq 1.18d'$ by (2.1). This is enough to find Q in R satisfying either (1), (2) or (3) (see Lemma 4.3). Note that this is not always possible if n is not bounded away from d , since then r may not be bounded away from d' . This is the reason we consider the ‘very dense’ case that $n < 1.19d$ separately.

To see how to guarantee (†), for the purposes of simplicity let us assume that Q satisfies (1), and further that Q has a subgraph C which is a cycle of length $c \geq 1.2d'$ (the remaining cases are similar but more involved, see Lemma 4.2). Write $C := 1 \dots c$. To show (†), we will use a variant of the Blow-up lemma [13, 17]. First, we do some pre-processing. Standard tools allow us to remove $o(|V_i|)$ vertices from each V_i with $i \in [c]$ to obtain new equally-sized clusters U_1, \dots, U_c , such that $G[U_i, U_{i+1}]$ is still dense and pseudorandom (where addition is modulo c), and also has large minimum degree. We say that $G' := \bigcup_{i \in [c]} G[U_i, U_{i+1}]$ is ‘super-regular with respect to C ’. Roughly speaking, a special case of the Blow-up lemma states that, for the purposes of embedding a (spanning) bounded degree subgraph, any graph H which is super-regular with respect to C behaves as if each $H[U_i, U_{i+1}]$ were *complete*.

For each $i \in [c]$, let $V'_i \subseteq U_i$ with $|V'_i| = m/2$ be arbitrary. Then one can show that, with high probability, the graph $H := \bigcup_{i \in [c]} G[V'_i, V'_{i+1}]$ is super-regular with respect to C . The Blow-up lemma tells us that, if the complete blow-up of the cycle C_c with parts of size $m/2$ contains a spanning cycle, then so does H . But this is clearly seen to be true (find the cycle by winding round parts). Thus almost every such V'_1, \dots, V'_c is such that $\bigcup_{i \in [c]} V'_i$ is a Hamiltonian subset. Moreover, the number of such choices is

$$(1 - o(1)) \cdot \prod_{i \in [c]} \binom{|U_i|}{m/2} \geq (1 - o(1)) \binom{(1 - o(1))m}{m/2}^{1.2d'} \geq 2^{1.1md'} \stackrel{(2.1)}{>} (2 - o(1))2^{d+1}.$$

So (†) holds, as required.

2.2. The sparse case: n/d is unbounded (Section 5). For the sparse case, we will work with expander graphs. This notion of expansion was first introduced by Komlós and Szemerédi [20]. Roughly speaking, a graph is an *expander* if every set that is not too large or small has large external neighbourhood (see Definition 5.1). The main property (Lemma 5.3) of expander graphs which we require is that

- (★) if H is an expander graph, then between every pair of large sets, there is a short path; and this path can be chosen to avoid an arbitrary small set.

Our aim in Section 5 is to prove the following (see Lemma 5.4).

- (‡) Given large $K > 0$, when d is sufficiently large and H is an almost d -regular expander graph on at least Kd vertices, then $c(H) \geq 2^{50d}$.

Here, almost d -regular means that there are constants $0 < \ell_1 < \ell_2$ such that $\ell_1 d \leq d_H(x) \leq \ell_2 d$ for all $x \in V(H)$. Komlós and Szemerédi [20] proved that every graph G contains an expander which has almost the same average degree as G (see Lemma 5.2). Then, roughly speaking, (‡) is used in our proof to ensure that any expander subgraph of G has at most Kd vertices, so we are in the dense case.

The proof of (‡) has the following general structure. We will locate a special set $Z \subseteq V(H)$ of size $200d$ with the following property (see Theorem 1.5 and Lemma 5.11). For every subset $U \subseteq Z$ of size $|Z|/2$, we can find a cycle C_U whose intersection with Z is almost the whole of U . This then implies that a large fraction of the $\binom{|Z|}{|Z|/2}$ cycles C_U are distinguishable.

To construct such a set Z , we will use different strategies depending on the edge density of H . In Section 5.3, we deal with the case when H is relatively dense. In this case, a key structure in our construction is a ‘web’ (See Definition 5.6 and Figure 3), which guarantees many vertex-disjoint paths within a relatively small neighbourhood of a ‘core’ vertex. This structure is inspired by some recent work on clique subdivisions (see [27] and [28]). We will iteratively construct many webs that are almost pairwise disjoint (in fact, they have disjoint ‘interiors’ (Lemma 5.7)), and let Z be the set of core vertices of these webs. Then for each $|Z|/2$ -set U in Z , to construct C_U , we will connect the webs corresponding to U via paths through their ‘exteriors’ in a cyclic manner. We hope to find vertex-disjoint (short) paths between the (large) exteriors of webs, avoiding previously-built paths, which together with the paths inside the webs leading to their core vertices form the desired cycle C_U . Property (★) is vital in obtaining these paths.

However, such ‘short’ paths can still block all webs, making it impossible to integrate their core vertices into the cycle C_U . To overcome this, we will choose our paths in a more careful way, such

that we avoid using too many vertices inside any particular web. Then the fact that the webs chosen are almost disjoint enables us to incorporate most of the vertices in U into C_U .

In Section 5.4, we deal with the case when H is very sparse. Here, an obvious difficulty in using the previous approach is that a single ‘short’ path could use the vertices in all webs due to the fact that the graph has very few edges. Instead, we will choose a set of vertices which are pairwise far apart in H to serve as Z . Such a set of vertices exists since H has small maximum degree compared to its order. To find the cycle C_U (see Lemma 5.11), we grow the neighbourhood around the vertices in U to a reasonably large size so that we can connect them via a path which avoids used vertices.

2.3. Finishing the proof (Section 6). There are several difficulties with the above approach. Statement (†) implies that any *almost d -regular* expander subgraph H of G must be dense. To find such an expander we must first remove large degree vertices from G and then apply Lemma 5.2 to find H . But then the average degree $d(H)$ of H could be less than d , in which case we may have $c(H) < c(K_{d+1})$ (for example if $H = K_d$). In the case when G is 2-connected, we circumvent this problem by finding two vertex-disjoint almost d -regular expander subgraphs H_1, H_2 of G and, for every fixed $x_i, y_i \in V(H_i)$, finding many x_i, y_i -paths P_i in H_i . The pairs (P_1, P_2) give rise to distinguishable cycles. (The general case is somewhat more technical and involves considering the so-called ‘block-structure’ of G and finding cycles and paths inside maximal 2-connected subgraphs of G .)

3. NOTATION AND PRELIMINARIES

3.1. Notation. For an integer N , let $[N] := \{1, \dots, N\}$. Given a set X and $k \in \mathbb{N}$, let $\binom{X}{k}$ be the collection of k -element subsets of X and let $\binom{X}{\leq k}$ be the collection of subsets of X with at most k elements. Define $\binom{X}{\geq k}$ analogously.

Given a graph G , we write $V(G)$ and $E(G)$ for its vertex and edge set respectively, and let $e(G) := |E(G)|$. Sometimes, we identify G with $V(G)$ by writing $x \in G$ instead of $x \in V(G)$ and $|G|$ instead of $|V(G)|$. For any $X \subseteq V(G)$, let $G - X$ be the graph obtained from G by removing the vertices of X and any edges incident to them. If $x \in V(G)$ we abbreviate $G - \{x\}$ to $G - x$. Given graphs G_1, G_2 , we write $G_1 \cup G_2$ to denote a graph with $V(G_1 \cup G_2) = V(G_1) \cup V(G_2)$ and $E(G_1 \cup G_2) = E(G_1) \cup E(G_2)$. Define the *neighbourhood* $N_G(X)$ and *external neighbourhood* $\Gamma_G(X)$ of X by setting

$$N_G(X) := \{u \in V(G) : uv \in E(G) \text{ for some } v \in X\} \quad \text{and} \quad \Gamma_G(X) := N_G(X) \setminus X.$$

Given $X \subseteq V(G)$ and $x \in V(G)$, write $d_G(x, X) := |N_G(x) \cap X|$. The *degree* of $x \in V(G)$ is $d_G(x) := d_G(x, V(G))$. Let $\delta(G)$ be the minimum degree of G and let $\delta_i(G) := d_i$ where $d_1 \leq d_2 \leq \dots \leq d_n$ is the degree sequence of the graph G . Write $\Delta(G)$ for the maximum degree and $d(G) := 2e(G)/|G|$ for the *average degree* of G .

We say that $W = (x_1, \dots, x_m)$ is a *walk* in a graph G if $x_i x_{i+1} \in E(G)$ for all $i \in [m-1]$; and a *circuit* if additionally $x_m x_1 \in E(G)$. For a given walk $W = (x_1, \dots, x_m)$ and vertex x , we let $\deg(x, W) = |\{i \in [m] : x = x_i\}|$. We say that a walk $P = (x_1, \dots, x_m)$ is a *path* if $x_i \neq x_j$ for all $i \neq j \in [m]$ and we define $\text{Int}(P) := \{x_2, \dots, x_{m-1}\}$ to be the *interior* of P . We say that a path $P = (x_1, \dots, x_m)$ is a u, v -*path* if $\{u, v\} = \{x_1, x_m\}$. The *length* of a path is the number of vertices it contains. We say a graph G *admits a vertex partition* (R, V_1, \dots, V_r) if $\{V_i : i \in [r]\}$ forms a partition of $V(G)$ into independent sets; R is a graph on vertex set $[r]$; and $G[V_i, V_j]$ is an empty graph for all $ij \notin E(R)$.

Unless otherwise specified, we write $\log(\cdot)$ for the natural logarithm $\log_e(\cdot)$. For $a, b, c \in \mathbb{R}$ we write $a = b \pm c$ if $b - c \leq a \leq b + c$. In order to simplify the presentation, we omit floors and ceilings and treat large numbers as integers whenever this does not affect the argument. The constants in the hierarchies used to state our results have to be chosen from right to left. More precisely, if we claim that a result holds whenever $0 < a \ll b \leq 1$, then this means that there is a non-decreasing function $f : (0, 1] \rightarrow (0, 1]$ such that the result holds for all $0 < a, b \leq 1$ with $a \leq f(b)$. Hierarchies with more constants are defined in a similar way.

Additional notation will be defined as and when it is required.

3.2. Tools. In this section we state some tools that we will use to prove our main result. For $n \in \mathbb{N}$ and $0 \leq p \leq 1$ we write $\text{Bin}(n, p)$ to denote the binomial distribution with parameters n and p . The hypergeometric random variable X with parameters (n, m, k) is defined as follows. We let N be a set of size n , fix $S \subseteq N$ of size $|S| = m$, pick uniformly at random a set $T \subseteq N$ of size $|T| = k$, then define $X = |T \cap S|$. Note that $\mathbb{E}X = km/n$. We will use the following standard Chernoff-type bound (see e.g. Theorem 2.10 in [11]).

Proposition 3.1. *Suppose X has binomial or hypergeometric distribution and $0 < a < 3\mathbb{E}[X]/2$. Then $\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq 2e^{-\frac{a^2}{3\mathbb{E}[X]}}$.*

The next result is well-known and guarantees that every graph has a subgraph with large average and minimum degrees.

Proposition 3.2. *Every graph G with average degree d contains a subgraph H with $d(H) \geq d$ and $\delta(H) \geq d/2$.*

We will also use the classical theorem of Erdős and Gallai from 1959 which gives a best-possible condition on the minimum length of a cycle in a graph with a given number of vertices and edges (we state a slightly weaker version here).

Theorem 3.3 ([10]). *For all $d \geq 2$, every graph G with $d(G) \geq d$ contains a cycle of length at least d .*

We will need the following theorem of Pósa, which gives a sufficient condition on the degree sequence of a graph for the presence of a Hamiltonian cycle. (Note that there are several strengthenings of this result, but the version below suffices for our purposes.)

Theorem 3.4 ([31]). *Let G be a graph on $n \geq 3$ vertices with degree sequence $d_1 \leq d_2 \leq \dots \leq d_n$. If $d_i \geq i + 1$ for all $i < (n - 1)/2$ and if additionally $d_{\lceil n/2 \rceil} \geq \lceil n/2 \rceil$ when n is odd, then G contains a Hamiltonian cycle.*

We will need the following easy bounds on certain binomial coefficients.

Proposition 3.5. *Suppose that $0 < 1/n \ll 1$ where $n \in \mathbb{N}$. Then*

$$\binom{n}{\lfloor n/4 - 1 \rfloor} \geq 2^{4n/5} \quad \text{and} \quad \binom{n}{2\lfloor n/4 - 1 \rfloor} \geq 2^{n - \log_2 n}.$$

Proof. Stirling's formula implies that, for all $m \geq 1$ and $k \geq 2$, we have

$$\sqrt{m} \binom{km}{m} \geq \frac{k^{k(m-1)+1}}{(k-1)^{(k-1)(m-1)}}.$$

So, setting $m := \lfloor n/4 - 1 \rfloor$, since $1/n \ll 1$, we have

$$\binom{n}{\lfloor n/4 - 1 \rfloor} \geq \binom{4m}{m} \geq \frac{4^{4m-3}}{\sqrt{m} \cdot 3^{3m-3}} = 2^{(8-3\log_2 3)m-6+3\log_2 3-(\log_2 m)/2} \geq 2^{3.24m} \geq 2^{4n/5}.$$

Now set $m' := 2\lfloor n/4 - 1 \rfloor$. Then

$$\binom{n}{2\lfloor n/4 - 1 \rfloor} \geq \binom{2m'}{m'} \geq 2^{2m'-1-(\log_2 m')/2} \geq 2^{n-\log_2 n},$$

as required. □

4. THE DENSE CASE

Our main aim in this section is, very roughly speaking, to prove Theorem 1.3 in the case when G is an n -vertex graph with average degree at least d and large minimum degree which is *dense* in the sense that d is linear in n . The two main results of this section are the following lemmas. Given distinct vertices $x, y \in V(G)$, denote by $p_{xy}(G)$ the number of vertex subsets U containing $\{x, y\}$ such that $G[U]$ contains a spanning x, y -path, which is precisely the number of distinguishable x, y -paths.

Lemma 4.1. *Suppose $0 < 1/d \ll \alpha \leq 1$ and that $n \in \mathbb{N}$ satisfies $d + 1 \leq n \leq 1.2d$. If G is an n -vertex graph with $d(G) \geq d$ and $\delta(G) \geq d/2$, then*

- (i) $c(G) \geq (1 - \alpha/2) \cdot 2^n$;
- (ii) if $x, y \in V(G)$ are distinct, then $p_{xy}(G) \geq (1 - \alpha/2) \cdot 2^{n-2}$.

Lemma 4.2. *Suppose $0 < 1/d \ll 1/K \leq 1$ and that $n \in \mathbb{N}$ satisfies $1.19d \leq n \leq Kd$. Let G be a 2-connected n -vertex graph with $d(G) \geq d$ and $\delta_2(G) \geq d/2$. If x, y are two distinct vertices of G , then*

$$c(G) > 2^{(1+\frac{1}{200})d} \quad \text{and} \quad p_{xy}(G) > 2^{0.89d}.$$

Lemma 4.1 addresses the case when G is almost complete, i.e. the number of vertices is at most $1.2d$. Its proof uses Theorem 3.4 and a probabilistic argument and appears in Section 4.1.

Lemma 4.2 addresses the case when our graph G on n vertices has average degree linear in n but also bounded away from n . In this case, we apply the Regularity lemma and use Lemma 4.3 to find in the reduced graph either (i) two vertex-disjoint cycles C_1, C_2 ; (ii) a sun S (see Section 4.2 for the definition); or (iii) a path P ; the clusters of which, in the original graph G , span at least $(1+1/100)d$ vertices. Then, using a probabilistic argument, we show that at least $2^{(1+1/200)d}$ subsets of the vertices lying in clusters of $C_1 \cup C_2$, S or P span a cycle. We use similar arguments to show that $p_{xy}(G)$ is large for any distinct $x, y \in V(G)$. The proof requires some more sophisticated tools and auxiliary results (Sections 4.2 to 4.3), so we defer its proof to Section 4.4.

4.1. The proof of Lemma 4.1. We do not need any additional tools to prove our first main lemma, and so we proceed immediately with the proof.

Proof of Lemma 4.1. Note that $\sum_{v \in V(G)} d_G(v) \geq dn$ and $d_G(v) \geq d/2$ for all $v \in V(G)$. Let

$$U := \{v \in V(G) : d_G(v) \geq 2d/3\}.$$

Then we get

$$dn \leq \sum_{v \in V(G)} d_G(v) \leq n|U| + 2d(n - |U|)/3,$$

and so, using the fact that $n \leq 1.2d$,

$$|U| \geq \frac{dn/3}{(n - 2d/3)} \geq \frac{5n}{8}.$$

To prove (i), choose a set $V' \subseteq V(G)$ uniformly at random by including each $v \in V'$ with probability $1/2$, independently of all other choices. Then for every $v \in V(G)$ we have

$$\mathbb{E}[|V'|] = n/2; \quad \mathbb{E}[|V' \cap U|] = |U|/2; \quad \text{and} \quad \mathbb{E}[d_G(v, V')] = d_G(v)/2.$$

Consider the following properties:

- (A1) $n' := |V'| = n/2 \pm d^{2/3}$;
- (A2) $|U \cap V'| \geq \frac{5n}{16} - d^{2/3}$;
- (A3) for all $v \in V(G)$ we have $d_G(v, V') \geq d_G(v)/2 - d^{2/3}$.

Then Proposition 3.1 implies that the probability that, when $1/d \ll \alpha$, all of (A1)–(A3) hold is at least

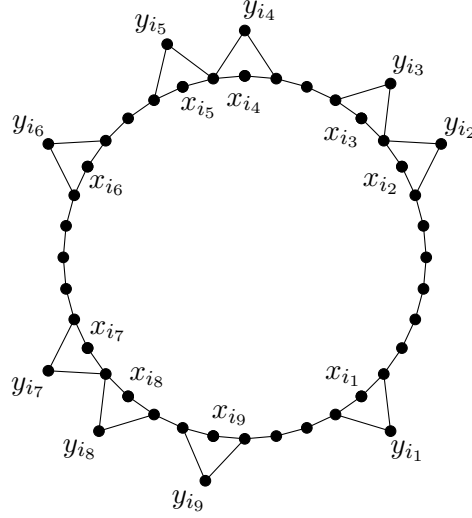
$$1 - \left(2e^{-\frac{2d^{4/3}}{3n}} + 2e^{-\frac{16d^{4/3}}{15n}} + \sum_{v \in V(G)} 2e^{-\frac{2d^{4/3}}{3d_G(v)}} \right) \geq 1 - e^{-d^{1/4}} \geq 1 - \frac{\alpha}{2},$$

where the penultimate inequality follows from $d_G(v) < n \leq 1.2d$ for all v .

Now, for a given set V' of size n' which satisfies (A1)–(A3), let $d_1 \leq \dots \leq d_{n'}$ be the degree sequence of the induced graph $G[V']$. We claim that this degree sequence satisfies the conditions of Theorem 3.4, and hence $G[V']$ is Hamiltonian. To see this, it is sufficient to find an integer k with $2 \leq k < n'/2$ such that $d_1 \geq k$ and $d_k > n'/2$. We claim that $k := 19n'/48$ suffices.

Indeed, (A1), (A3) and the fact that $\delta(G) \geq d/2$, $n \leq 1.2d$ and $1/d \ll 1$ imply that

$$d_1 \geq \frac{d}{4} - d^{2/3} \geq \frac{5n}{24} - d^{2/3} \geq \frac{5n'}{12} - 2d^{2/3} > k.$$

FIGURE 2. A $(36, 9)$ -sun with labelled corona.

Also, (A1) and (A2) imply that $|U \cap V'| \geq 5n'/8 - 2d^{2/3}$, and (A3) together with the definition of U implies that, for all $v \in U \cap V'$, we have

$$d_G(v, V') \geq \frac{d}{3} - d^{2/3} \geq \frac{5n'}{9} - 2d^{2/3} > \frac{n'}{2}.$$

Thus at least

$$|U \cap V'| \geq \frac{5n'}{8} - 2d^{2/3} > \frac{29n'}{48} + 1 = n' - k + 1$$

vertices in V' have degree at least $n'/2$, and so $d_k > n'/2$, as required.

We have shown that, for at least $(1 - \alpha/2) \cdot 2^n$ distinct subsets $V' \subseteq V(G)$, we have that $G[V']$ is Hamiltonian. Thus $c(G) \geq (1 - \alpha/2) \cdot 2^n$, proving (i).

For part (ii), the argument is very similar. Choose a set $V' \subseteq V(G)$ uniformly at random among all sets V' containing both x and y (i.e. include each of x, y with probability one and every other vertex v with probability $1/2$, independently of all other choices). A near identical argument shows that (A1)–(A3) hold with probability at least $1 - \alpha/2$. Now obtain G' from $G[V']$ by adding a new vertex z and the edges xz, yz . We now have that

$$d_1 = d_{G'}(z) = 2; \quad d_2 \geq k; \quad \text{and} \quad d_k > n'/2.$$

Theorem 3.4 implies that G' contains a Hamiltonian cycle C . Since $d_{G'}(z) = 2$, C contains the edges xz, yz . Thus C without these edges is a path with endpoints x, y which spans V' . There are 2^{n-2} subsets of $V(G)$ containing both x and y ; so for at least $(1 - \alpha/2) \cdot 2^{n-2}$ choices of V' , we have such a path. Thus $p_{xy}(G) \geq (1 - \alpha/2) \cdot 2^{n-2}$, proving (ii). \square

4.2. Suns, paths and cycles. Let $a, b \in \mathbb{N}$ and consider a sequence $I = (i_1, \dots, i_b)$ with $1 \leq i_1 < i_2 < \dots < i_b \leq a$ such that $i_j - i_{j-1} \geq 2$ and $i_1 + a - i_b \geq 2$. Define a graph S as follows. Let $V(S) := \{x_1, \dots, x_a\} \cup \{y_i : i \in I\}$, $x_{a+1} := x_1$ and $x_0 := x_a$. Let

$$E(S) := \{x_i x_{i+1} : i \in [a]\} \cup \{x_{i-1} y_i, y_i x_{i+1} : i \in I\}.$$

We call such a graph S an (a, b) -sun. Note that by definition $b \leq a/2$. The set of vertices remaining after deleting $\{x_i : i \in I\}$ from S induce a cycle in S ; and similarly after deleting $\{y_i : i \in I\}$. We call $\text{Cor}(S) := \{x_i, y_i : i \in I\}$ the *corona* of S . See Figure 2 for an illustration of a sun.

In the next lemma, we show that any graph G with average degree at least d and minimum degree at least $d/2$, and whose vertex set has size at least slightly larger than d , contains either (i) two vertex-disjoint cycles, the sum of whose lengths is large; (ii) a long path; or (iii) a large sun with large corona. Later we will use Lemma 4.3 to find these structures in a reduced graph obtained after applying the Regularity lemma.

Lemma 4.3. *Let $n \geq 1.18d$ and let G be an n -vertex graph with $d(G) \geq d$ and $\delta(G) \geq d/2$. Then G contains at least one of the following:*

- (i) two vertex-disjoint cycles C_1, C_2 with $|C_1| + |C_2| \geq 1.8d$;
- (ii) a path P with $|P| \geq (1 + 1/100)d$;
- (iii) an (a, b) sun with $a \geq d$ and $b \geq d/20$.

Proof. First, we show that either there are two disjoint cycles in G satisfying (i), or there is a large dense component H in G . In the latter case, we will show that H contains the configuration in (ii) or (iii).

By averaging, we can find in G at least one component H with $d(H) \geq d(G) \geq d$. Let $n_1 = |V(H)|$, $d_1 = d(H)$, and $d_2 = d(G \setminus H)$. Observe that we may assume $d_2 < d$, since otherwise Theorem 3.3 guarantees two vertex-disjoint cycles, one in H and one in $G \setminus H$, with total length at least $d_1 + d_2 \geq 2d$, implying (i). We claim that we may additionally assume

$$d \leq d(G) \leq d_1 \leq 1.01d \quad \text{and} \quad \frac{n}{n_1} < \frac{22}{21}. \quad (4.1)$$

Indeed, if $d_1 \geq 1.01d$, then by Theorem 3.3 there is a cycle of length at least $1.01d$, implying (ii). To see why we may assume the second inequality in (4.1), note that as $d - d_2 > 0$, we have

$$n_1 d_1 + (n - n_1) d_2 \geq nd \quad \Leftrightarrow \quad n_1(d_1 - d_2) \geq n(d - d_2) \quad \Leftrightarrow \quad \frac{n}{n_1} \leq \frac{d_1 - d_2}{d - d_2}. \quad (4.2)$$

Now, if $n/n_1 \geq 22/21$ then by (4.2), we have

$$\frac{d_1 - d_2}{d - d_2} \geq \frac{22}{21} \quad \Rightarrow \quad d_1 + d_2 \geq 22d - 20d_1 \stackrel{(4.1)}{\geq} 22d - 20.2d = 1.8d.$$

By Theorem 3.3, we can find in H a cycle C_1 of length at least d_1 and in $G \setminus H$ a cycle C_2 of length at least d_2 . The above inequality implies $|C_1| + |C_2| \geq d_1 + d_2 \geq 1.8d$, yielding (i).

From now on, we will work in the connected n_1 -vertex graph H which has average degree d_1 . Then (4.1) together with $n \geq 1.18d$ implies that

$$\delta(H) \geq \frac{d}{2} \geq \frac{d_1/1.01}{2} > 0.495d_1 \quad \text{and} \quad n_1 > \frac{21}{22}n \geq \frac{21}{22} \cdot 1.18 \cdot \frac{d_1}{1.01} \geq 1.115d_1. \quad (4.3)$$

Let $C = x_1 \dots x_a$ be a longest cycle in H and set $x_{a+1} := x_1$. Theorem 3.3 implies that $a \geq d_1$. Again, $a \leq 1.01d_1$, since otherwise (ii) holds by the fact that $d_1 \geq d$. Now let $U := V(H) \setminus V(C)$. Then, using (4.3),

$$|U| = n_1 - a \geq 1.115d_1 - 1.01d_1 \geq 0.105d_1. \quad (4.4)$$

Suppose that $d(H[U]) \geq d_1/100$. By Theorem 3.3, $H[U]$ contains a cycle C' with $|C'| \geq d_1/100$. Since H is connected, there is a path between the vertex-disjoint cycles C and C' . This path together with C and C' forms a path of length at least $|C| + |C'| \geq 1.01d_1$, and so (ii) holds.

Thus we may assume that $d(H[U]) < d_1/100$. If there are at most $d_1/20$ vertices $u \in U$ such that $d_H(u, U) \leq d_1/40$, then, by (4.4)

$$d(H[U])|U| = \sum_{u \in U} d_H(u, U) \geq \left(|U| - \frac{d_1}{20} \right) \frac{d_1}{40} > \frac{d_1}{100} \cdot |U|,$$

a contradiction. Thus, writing $b := d_1/20$, we can find distinct vertices z_1, \dots, z_b in U such that $d_H(z_j, U) \leq d_1/40$ for all $j \in [b]$. Since $\delta(H) > 0.495d_1$, for every $j \in [b]$,

$$d_H(z_j, C) \geq \delta(H) - d_H(z_j, U) > 0.495d_1 - \frac{d_1}{40} = 0.47d_1. \quad (4.5)$$

We will now iteratively construct a sequence of distinct indices i_1, \dots, i_b in $[a]$ such that the following hold for all $j \in [b]$:

- (S1) $_j$ $x_{i_j-1}, x_{i_j+1} \in N_H(z_j)$;
- (S2) $_j$ $|i_j - i_\ell| \not\equiv 0, 1 \pmod{a}$ for all $\ell \in [j-1]$.

Suppose that, for some $j \leq b$, we have chosen distinct i_1, \dots, i_{j-1} in $[a]$ such that (S1) $_{j'}$ and (S2) $_{j'}$ hold for all $j' \in [j-1]$. We will show that we can choose an appropriate i_j . To this end, we use (4.5) to choose $1 \leq k_1 < \dots < k_{0.47d_1} \leq a$ so that x_{k_i} is a neighbour of z_j for all $i \in [0.47d_1]$. For each $t \in \mathbb{N}$, let p_t be the number of indices $\ell \in [0.47d_1]$ such that $k_{\ell+1} - k_\ell = t$ if $\ell < 0.47d_1$; or $k_1 - k_{0.47d_1} = t - a$. Notice first that $p_1 = 0$, otherwise z_j has a pair of consecutive neighbours in

C , contradicting to the maximality of C . By definition, p_2 is at most the number of indices $i \in [a]$ for which $x_{i-1}, x_{i+1} \in N_H(z_j)$, i.e. which satisfy $(S1)_j$. Then, since $\sum_t p_t = 0.47d_1$,

$$1.01d_1 \geq a = \sum_{t \in \mathbb{N}} t \cdot p_t \geq 2p_2 + 3 \sum_{t \geq 3} p_t \geq 2p_2 + 3(0.47d_1 - p_2) = 1.41d_1 - p_2,$$

and so $p_2 \geq 0.4d_1$. Thus there are at least $0.4d_1$ indices i in $[a]$ which satisfy $(S1)_j$. Now, the number of indices $i \in [a]$ which do not satisfy $(S2)_j$ is at most $3(j-1)$ (the indices i_1, \dots, i_{j-1} already obtained, as well as their neighbours on C). But

$$3(j-1) < 3b = \frac{3d_1}{20} < 0.4d_1 \leq p_2.$$

Therefore, we can choose $i_j \in [a]$ such that $(S1)_{j'}$ and $(S2)_{j'}$ hold for all $j' \in [j]$ and obtain i_1, \dots, i_b as required. We have found an (a, b) -sun with vertex set $C \cup \{z_1, \dots, z_b\}$ and corona $\{x_{i_j} : j \in [b]\} \cup \{z_1, \dots, z_b\}$. So (iii) holds, completing the proof. \square

The next proposition shows that, for any cycle, sun or path H , there is a walk W in H such that the set of values $\deg(x, W)$ for $x \in V(H)$ is well-controlled. Its proof is straightforward, we defer it to the appendix.

Proposition 4.4. *Let $n, a, b \in \mathbb{N}$ where $b \leq a/2$. Let H be a graph and let $u, v \in V(H)$.¹ Suppose that H is either (i) a cycle C on a vertices; (ii) a path Q on a vertices; or (iii) an (a, b) -sun S . Then in each case, there exists a walk W in H from u to v such that*

- (i) $N_2 = V(C)$.
- (ii) N_1 is the set of two endpoints of Q and N_2 is the set of internal vertices of Q .
- (iii) $N_1 = \text{Cor}(S)$ and $N_2 := V(S) \setminus \text{Cor}(S)$;

where

$$N_k := \{x \in V(H) : \deg(x, W) \in \{kn, kn+1, kn+2\}\}.$$

4.3. The Regularity and Blow-up lemmas. In our proof, we apply Szemerédi's Regularity lemma [32]. For a comprehensive survey of this lemma and its many applications, see [18, 19]. To state the lemma we need some more definitions. We write $d_G(A, B)$ for the *density* $\frac{e(G[A, B])}{|A||B|}$ of a bipartite graph G with vertex classes A and B . Given $\varepsilon > 0$, we say that G is ε -regular if every $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ satisfy $|d(A, B) - d(X, Y)| \leq \varepsilon$. Given $\varepsilon, \gamma \in (0, 1)$ we say that G is (ε, γ) -regular if G is ε -regular and $d_G(A, B) \geq \gamma$. We say that G is (ε, γ) -super-regular if both of the following hold:

- G is (ε, γ) -regular;
- $d_G(a) \geq \gamma|B|$ and $d_G(b) \geq \gamma|A|$ for all $a \in A, b \in B$.

Let R be a graph on vertex set $[r]$. We say that a graph H is (ε, γ) -(super-)regular with respect to vertex partition (R, V_1, \dots, V_r) if $\{V_i : i \in V(R)\}$ is a partition of $V(H)$; and $H[V_i, V_j]$ is (ε, γ) -(super-)regular for all $ij \in E(R)$.

We will use the following degree form of the Regularity lemma, which can be easily derived from the usual version (see [19]).

Lemma 4.5. (Degree form of the Regularity lemma) *Suppose $0 < 1/n \ll 1/M \ll \varepsilon, 1/M' < 1$ with $n \in \mathbb{N}$ and $\gamma \in \mathbb{R}$. Then there is a partition of the vertex set of G into V_0, V_1, \dots, V_r and a spanning subgraph G' of G such that the following holds:*

- (i) $M' \leq r \leq M$;
- (ii) $|V_0| \leq \varepsilon n$;
- (iii) $|V_1| = \dots = |V_r|$;
- (iv) $d_{G'}(x) > d_G(x) - (\gamma + \varepsilon)n$ for all $x \in V(G)$;
- (v) for all $i \in [r]$ the graph $G'[V_i]$ is empty;
- (vi) for all $1 \leq i < j \leq r$, $G'[V_i, V_j]$ is ε -regular and has density either 0 or at least γ .

We call V_1, \dots, V_r *clusters* and refer to G' as the *pure graph*. The last condition of the lemma says that all pairs of clusters are ε -regular (but possibly with different densities). The *reduced graph*

¹When not specified, u and v need not be distinct.

R of G with parameters ε , γ and M' is the graph whose vertex set is $[r]$ and in which ij is an edge precisely when $G'[V_i, V_j]$ is ε -regular and has density at least γ .

The following is a well-known tool which is similar to Proposition 8 in [25], which states that a regular partition of a graph can be made into a super-regular partition by a slight modification.

Lemma 4.6 ([25]). *Suppose $0 < 1/m \ll \varepsilon \ll \gamma, 1/\Delta_R \leq 1$ and $1/m \ll 1/r$ with $r, m \in \mathbb{N}$. Let R be a graph with $V(R) = [r]$ and $\Delta(R) \leq \Delta_R$. Let G be a (ε, γ) -regular graph with respect to vertex partition (R, V_1, \dots, V_r) such that $|V_i| = (1 \pm \varepsilon)m$ for all $1 \leq i \leq r$. Then there exists $G' \subseteq G$ which is $(4\sqrt{\varepsilon}, \gamma/2)$ -super-regular with respect to vertex partition (R, V'_1, \dots, V'_r) such that $V'_i \subseteq V_i$ and $|V'_i| = (1 - \sqrt{\varepsilon})m$.*

The following ‘slicing’ lemma states that every induced subgraph of a regular pair which is not too small is still regular (with a weaker regularity parameter).

Lemma 4.7. *Let $\varepsilon, \gamma \in \mathbb{R}$ be such that $\gamma > 2\varepsilon > 0$. Let (A, B) be an (ε, γ) -regular pair. Suppose that $A' \subseteq A$ and $B' \subseteq B$ are such that $|A'| \geq \sqrt{\varepsilon}|A|$ and $|B'| \geq \sqrt{\varepsilon}|B|$. Then (A', B') is an $(\sqrt{\varepsilon}, \gamma/2)$ -regular pair.*

For technical reasons, we will use a modification of the celebrated Blow-up lemma of Komlós, Sárközy and Szemerédi (see Remark 8 in [17]). Indeed, this result does not allow the situation in which $1/r$ is much smaller than ε , as we require. Although Csaba proved a version which does allow this situation (see Lemma 5 in [6]), his result is not formulated to allow a non-equitable partition. Thus we use the following lemma from [13] (a simplified version of Theorem 6 in [13]).

The Blow-up lemma states that, for the purpose of embedding a (spanning) bounded degree graph H , a graph G which is super-regular with respect to some graph J behaves like a complete ‘blow-up’ of J . Moreover, if there are a small number of special vertices in H which each have a large ‘target set’ in G , then there is an embedding of H that maps every special vertex into its associated target set.

Theorem 4.8. *Suppose $0 < 1/n \ll \varepsilon, \alpha \ll \gamma, 1/\Delta, 1/\Delta_R, c \leq 1$ and $1/n \ll 1/r$ where $r, n \in \mathbb{N}$. Suppose further that*

- (i) $n \leq n_i \leq n + 2$, for all $i \in [r]$;
- (ii) R is a graph on vertex set $[r]$ with $\Delta(R) \leq \Delta_R$;
- (iii) G is (ε, γ) -super-regular with respect to vertex partition (R, V_1, \dots, V_r) such that $|V_i| = n_i$ for all $i \in [r]$;
- (iv) H is a graph which admits a vertex partition (R, X_1, \dots, X_r) such that $\Delta(H) \leq \Delta$ and $|X_i| = n_i$.

Moreover, suppose that in each class X_i there is a set of αn_i special vertices $y \in V(H)$, each of them equipped with a set $S_y \subseteq V_i$ with $|S_y| \geq cn_i$. Then there is an embedding of H into G such that every special vertex y is mapped to a vertex in S_y .

We will use Theorem 4.8 to prove the following lemma which is essentially the special case when H is a path, except that we now prescribe the endpoints of the embedding of H in G (rather than prescribing a set of allowed endpoints) and some sets V_i are twice the size of others. Again, the proof is deferred to the appendix.

Lemma 4.9. *Suppose $0 < 1/n \ll \varepsilon \ll \gamma, 1/\Delta_R \leq 1$ and $1/n \ll 1/r$ where $r, n \in \mathbb{N}$. Suppose further that*

- (i) $n_i \in \{n, n + 1, n + 2, 2n, 2n + 1, 2n + 2\}$ for all $i \in [r]$;
- (ii) R is a graph on vertex set $[r]$ with $\Delta(R) \leq \Delta_R$;
- (iii) G is a graph and $x, y \in V(G)$ are two distinct vertices. Suppose further that G is (ε, γ) -super-regular with respect to vertex partition (R, V_1, \dots, V_r) , where $x \in V_j$ and $y \in V_k$; and $|V_i| = n_i$ for all $i \in [r]$.
- (iv) $W = (w_1, \dots, w_m)$ is a walk in R from j to k such that $\deg(i, W) = n_i$ for all $i \in [r]$.

Then G contains a spanning path P with endpoints x, y .

4.4. The proof of Lemma 4.2. Let $K \geq 1.19$ be given. Choose integers M, M', d_0 and constants ε, γ such that

$$0 < 1/d_0 \ll 1/M \ll 1/M' \ll \varepsilon \ll \gamma \ll 1/K.$$

Now fix $d \geq d_0$. Let G be a 2-connected graph on n vertices such that $1.19d \leq n \leq Kd$, $d(G) \geq d$ and $\delta_2(G) \geq d/2$.

Apply the Regularity lemma (Lemma 4.5) to G with parameters ε, γ, M' to obtain clusters V_1, \dots, V_r of equal size, an exceptional set V_0 of size at most εn , and a pure graph G' which is (ε, γ) -regular with respect to a vertex partition (R, V_1, \dots, V_r) . The vertex set of R is $[r]$, where $M' \leq r \leq M$. Let $m := |V_1| = \dots = |V_r|$ and $\xi := d/n$. Then

$$1/K \leq \xi \leq 100/119 \quad \text{and} \quad (1 - \varepsilon)n \leq rm \leq n. \quad (4.6)$$

First we show that R essentially inherits the minimum and average degree of G .

Claim 4.10. *Let $d' := (1 - \sqrt{\gamma})\xi r$. Then*

$$r \geq 1.18d'; \quad \delta(R) \geq d'/2; \quad \text{and} \quad d(R) \geq d'.$$

Proof: The first inequality is immediate from the definition of d' , that $\gamma \ll 1/K$ and that $\xi \leq 100/119$. For a contradiction to the second, suppose there is some $i \in [r]$ with $d_R(i) < d'/2$. Count the number of edges e_i in G' incident to V_i as follows. Since $\delta_2(G) \geq \xi n/2$, we have, using Lemma 4.5(iv) and (v), that

$$e_i = \sum_{v \in V_i} d_{G'}(v) \geq (\xi/2 - \varepsilon - \gamma)n(m-1) \geq (\xi/2 - 2\gamma)nm. \quad (4.7)$$

By assumption, less than $d'/2$ clusters V_j are such that (V_i, V_j) is an (ε, γ) -regular pair in G' , and $G'[V_i, V_j]$ is empty for all other V_j . So

$$e_i < d'm^2/2 \leq (1 - \sqrt{\gamma})\xi nm/2,$$

since $rm \leq n$. Together with (4.7), this gives a contradiction because $2\gamma < \xi\sqrt{\gamma}/2$. This proves that $\delta(R) \geq d'/2$.

For the final inequality, since $rm \leq n \leq Kd$ and $\varepsilon \ll \gamma \ll 1/K$,

$$\begin{aligned} d(R) &= \frac{2e(R)}{r} \geq \frac{2e(G')}{rm^2} \geq \frac{(d - (\gamma + \varepsilon)n)n}{rm^2} \geq \frac{d - 2\gamma \cdot Kd}{m} \stackrel{(4.6)}{\geq} \frac{(1 - 2\gamma K)dr}{n} \\ &\geq (1 - \sqrt{\gamma})\frac{dr}{n} = d', \end{aligned}$$

as required. ■

Claim 4.10 together with Lemma 4.3 implies that R contains one of the following:

- (1) vertex-disjoint cycles H, J , where $|J| \leq |H| \leq 101d'/100$ and $|H| + |J| \geq 1.8d'$;
- (2) a path H with $|H| = (1 + 1/100)d'$;
- (3) an (a, b) sun H with $d' \leq a < 101d'/100$ and $d'/20 \leq b \leq a/2$.

(If the upper bounds in (1) and (3) do not hold, then we have an instance of (2).) Observe that, in all cases, $\Delta(H) \leq 4$ and $|H| \leq 2d'$. Throughout the rest of this proof, denote

$$m' := (1 - \sqrt{\varepsilon})m \quad \text{and} \quad m_0 := \lfloor m'/4 - 1 \rfloor. \quad (4.8)$$

Lemma 4.6 with H and G' playing the roles of R and G respectively implies that, for each $i \in V(H)$, V_i contains a subset V'_i of size m' such that for every edge ij of H , the graph $G'[V'_i, V'_j]$ is $(4\sqrt{\varepsilon}, \gamma/2)$ -super-regular. Let $V' := \bigcup_{i \in V(H)} V'_i$. Let $x', y' \in V'$ be distinct vertices and $i_1, i_2 \in V(H)$ be such that $x' \in V'_{i_1}$ and $y' \in V'_{i_2}$. Apply Proposition 4.4 with m_0, i_1 and i_2 playing the roles of n, u and v respectively to obtain a partition $N_1 \cup N_2$ of $V(H)$ such that there is a walk $W = (x_1, \dots, x_\ell)$ in H such that $x_1 = i_1$ and $x_\ell = i_2$, where for each $k \in [2]$,

$$N_k := \{i \in [r] : \deg(i, W) \in \{km_0, km_0 + 1, km_0 + 2\}\}. \quad (4.9)$$

Let $n_i := \deg(i, W)$ for all $i \in V(H)$. Note that $n_i \in \{m_0, m_0 + 1, m_0 + 2, 2m_0, 2m_0 + 1, 2m_0 + 2\}$. So a crude but useful estimate is that, for all $i \in V(H)$,

$$n_i \geq m'/5. \quad (4.10)$$

Claim 4.11. *If $4|N_1|/5 + |N_2| \geq \gamma d'$, then*

$$p_{x'y'}(G'[V']) \geq 2^{(1-2\sqrt{\gamma})(4|N_1|/5 + |N_2|)d/d'}.$$

Proof: Suppose that $X \subseteq V(G)$ is such that

- (i) $X = \bigcup_{i \in V(H)} X_i$ where $X_i \subseteq V'_i$ and $|X_i| = n_i$;
- (ii) $x' \in X_{i_1}$ and $y' \in X_{i_2}$;
- (iii) $G'[X_i, X_j]$ is $(\varepsilon^{1/3}, \gamma/3)$ -super-regular for all $ij \in E(H)$.

Then Lemma 4.9 applied with the following graphs and parameters implies that $G'[X]$ contains a spanning path with endpoints x' and y' :

| object/parameter | H | $G'[X]$ | X_i | W | $\varepsilon^{1/3}$ | $\gamma/3$ | x' | y' | i_1 | i_2 | m_0 |
|---------------------|-----|---------|-------|-----|---------------------|------------|------|------|-------|-------|-------|
| playing the role of | R | G | V_i | W | ε | γ | x | y | j | k | n . |

Thus $p_{x'y'}(G'[V'])$ is at least the number of subsets X of $V(G)$ satisfying (i)–(iii). We claim that at least half of the sets X satisfying the first two properties also satisfy the third. Indeed, for each $i \in V(H)$, independently choose $X_i \subseteq V'_i$ uniformly at random among all subsets X_i of V'_i such that $x \in X_{i_1}$ and $y \in X_{i_2}$ and $|X_i| = n_i$. (So $X := \bigcup_{i \in V(H)} X_i$ satisfies (i) and (ii).) Now fix $z \in V'_i$ and $j \in N_H(i)$. Then, since $G'[V'_i, V'_j]$ is $(4\sqrt{\varepsilon}, \gamma/2)$ -super-regular,

$$n_j = |X_j| \geq \mathbb{E}[d_{G'}(z, X_j)] \geq d_{G'}(z, V'_j) \cdot \frac{n_j - 2}{m'} \geq \frac{\gamma m'}{2} \cdot \frac{n_j}{m'} - 2 \geq \frac{5\gamma n_j}{12}.$$

So Lemma 3.1 implies that

$$\mathbb{P}\left[d_{G'}(z, X_j) < \frac{\gamma n_j}{3}\right] \leq 2e^{-\frac{(\gamma n_j/12)^2}{3n_j}} \leq 2e^{-\gamma^3 n_j} \stackrel{(4.10)}{\leq} e^{-\sqrt{m'}}.$$

So the probability that $d_{G'}(z, X_j) \geq \gamma n_j/3$ for every $z \in V'_i$ with $i \in V(H)$ and every $j \in N_H(i)$ is at least

$$1 - |H| \cdot m' \cdot \Delta(H) \cdot e^{-\sqrt{m'}} \geq 1 - 2n^2 \cdot 4 \cdot e^{-(n/r)^{1/3}} \geq \frac{1}{2}$$

since $1/n \ll 1/r$. We have shown that at least half of the $X \subseteq V(G)$ satisfying both (i) and (ii) are such that $d_{G'}(x, X_j) \geq \gamma n_j/3$ for all $i \in V(H)$, $x \in X_i$ and $j \in N_H(i)$. Call such a set X *good*.

Lemma 4.7 together with (4.8), (4.10) and that $G'[V_i, V_j]$ is (ε, γ) -regular implies that for all good X we have that $G'[X_i, X_j]$ is $(\sqrt{\varepsilon}, \gamma/2)$ -regular for all $ij \in E(H)$, and hence $(\varepsilon^{1/3}, \gamma/3)$ -regular. Therefore, since X is good, each such $G'[X_i, X_j]$ is $(\varepsilon^{1/3}, \gamma/3)$ -super-regular. Thus every good X automatically satisfies (iii) and the number of X satisfying (i)–(iii) is at least the number of good X .

When $i_1 = i_2 =: i^*$, this shows that

$$\begin{aligned} p_{x'y'}(G'[V']) &\geq \frac{1}{2} \cdot \binom{m' - 2}{n_{i^*} - 2} \prod_{i \in V(H) \setminus \{i^*\}} \binom{m'}{n_i} = \frac{1}{2} \cdot \frac{n_{i^*}(n_{i^*} - 1)}{m'(m' - 1)} \prod_{i \in V(H)} \binom{m'}{n_i} \\ &\stackrel{(4.10)}{>} \frac{1}{100} \prod_{i \in V(H)} \binom{m'}{n_i}. \end{aligned}$$

The same bound holds when $i_1 \neq i_2$. Now, by (4.8) and (4.9), $n_i \leq m'/2$ for all $i \in N_1 \cup N_2$. So $\binom{m'}{n_i} \geq \binom{m'}{2m_0}$ for all $i \in N_2$ and $\binom{m'}{n_i} \geq \binom{m'}{m_0}$ for all $i \in N_1$. By Proposition 3.5, we have

$$\begin{aligned} p_{x'y'}(G[V']) &\geq \frac{1}{100} \prod_{i \in N_1} \binom{m'}{\lfloor m'/4 - 1 \rfloor} \prod_{j \in N_2} \binom{m'}{2\lfloor m'/4 - 1 \rfloor} \\ &\geq \frac{1}{100} \cdot 2^{(4|N_1|/5 + |N_2|)m' - |N_2|\log_2 m'} \geq \frac{1}{100} \cdot 2^{(1-\gamma)(4|N_1|/5 + |N_2|)m'}, \end{aligned} \quad (4.11)$$

where, for the final inequality, we used the fact that $1/m' \ll \gamma$. Suppose that $4|N_1|/5 + |N_2| \geq \gamma d'$. Then using the fact that $\frac{n}{r} = (1 - \sqrt{\gamma})\frac{d}{d'}$, we have

$$\begin{aligned} (1 - \gamma) \left(\frac{4|N_1|}{5} + |N_2| \right) m' &\stackrel{(4.6), (4.8)}{\geq} (1 - \gamma) \left(\frac{4|N_1|}{5} + |N_2| \right) \frac{(1 - \sqrt{\varepsilon})(1 - \varepsilon)n}{r} \\ &\geq (1 - 2\gamma) \left(\frac{4|N_1|}{5} + |N_2| \right) (1 - \sqrt{\gamma}) \frac{d}{d'} \\ &\geq (1 - 2\sqrt{\gamma}) \left(\frac{4|N_1|}{5} + |N_2| \right) \frac{d}{d'} + 10. \end{aligned}$$

We get the last inequality by the facts $1/d \ll \gamma$ and $4|N_1|/5 + |N_2| \geq \gamma d'$. Together with (4.11), this proves Claim 4.11. \blacksquare

Now let x, y be arbitrary distinct vertices of $V(G)$. Since G is 2-connected, there exist two vertex-disjoint paths from $\{x, y\}$ to V' . Let P_x, P_y be two such minimal paths, where P_x is from x to some $x'' \in V'$; and P_y is from y to some $y'' \in V'$. As the choices of $x', y' \in V'$ were arbitrary, by letting $x'' = x'$ and $y'' = y'$, we see that every distinguishable x', y' -path in $G[V']$ together with P_x and P_y forms a distinguishable x, y -path in G , namely we have

$$p_{xy}(G) \geq p_{x'y'}(G'[V']). \quad (4.12)$$

We now return to cases (1)–(3) to prove the assertion concerning $p_{xy}(G)$.

Case (1): H is a cycle with $0.9d' \leq |H| \leq 101d'/100$.

Note that the lower bound on $|H|$ is implied by the existence of J in this case and that $|H| \geq |J|$, $|H| + |J| \geq 1.8d'$. Now Proposition 4.4 implies that $|N_2| = |H| \geq 0.9d'$, so $4|N_1|/5 + |N_2| \geq 0.9d'$. Thus (4.12) and Claim 4.11 with the fact that $\gamma < 10^{-5}$ imply that

$$p_{xy}(G) \geq p_{x'y'}(G'[V']) \geq 2^{(1-2\sqrt{\gamma})|H| \cdot d/d'} \geq 2^{0.89d}, \quad (4.13)$$

as required.

Case (2): H is a path on $a \geq 101d'/100$ vertices.

Now $|N_2| = |H| - 2$, so $4|N_1|/5 + |N_2| \geq 101d'/100 - 2$. Thus, Claim 4.11 implies that

$$p_{xy}(G) \geq 2^{(1+\frac{1}{150})d}.$$

Case (3): H is an (a, b) -sun with $a \geq d'$ and $b \geq d'/20$.

Now $|N_1| = |\text{Cor}(H)| = 2b$ and $|N_2| = a - b$. So

$$\frac{4|N_1|}{5} + |N_2| = \frac{3b}{5} + a \geq \left(1 + \frac{3}{100}\right)d'.$$

Similarly by Claim 4.11, we have that $p_{xy}(G) \geq 2^{(1+\frac{1}{150})d}$.

This completes the proof that $p_{xy}(G) \geq 2^{0.89d}$. Proving the assertion about $c(G)$ is now easy in Cases (2) and (3). Here, for arbitrary distinct vertices x, y , we have that $p_{xy}(G) \geq 2^{(1+1/150)d}$. Now choose x, y such that $xy \in E(G)$. At most one x, y -path uses this edge and so we have that

$$c(G) \geq p_{xy}(G) - 1 \geq 2^{(1+1/200)d}.$$

Therefore we may assume that we are in Case (1). Using exactly the same arguments as above for J instead of H , we can obtain a set $U' \subseteq V(G)$ (the analogue of V'), which is a subset of $\bigcup_{i \in V(J)} V_i$, such that for arbitrary distinct vertices $u', v' \in U'$, we have $p_{u'v'}(G'[U']) \geq 2^{(1-2\sqrt{\gamma})|J| \cdot d/d'}$ (in analogy with the middle inequality in (4.13)).

Observe that, because H, J are vertex-disjoint subgraphs of R , the sets U', V' are disjoint. Since G is 2-connected, there exist two vertex-disjoint paths between U' and V' . Choose minimal such paths P_1 with endpoints x', u' and P_2 with endpoints y', v' , where $x', y' \in V'$ and $u', v' \in U'$. Then distinguishable x', y' -paths in V' and u', v' -paths in U' together with P_1, P_2 yield distinct Hamiltonian subsets in G . Recall that $|H| + |J| \geq 1.8d'$. We then have

$$c(G) \geq p_{x'y'}(G'[V']) \cdot p_{u'v'}(G'[U']) \geq 2^{(1-2\sqrt{\gamma})(|H|+|J|) \cdot d/d'} \geq 2^{1.7d},$$

as required. This completes the proof of Lemma 4.2. \square

5. THE SPARSE CASE

In this section, we will prove the second main ingredient, which states that a large almost-regular expander graph contains many Hamiltonian subsets, see Lemma 5.4. To state it formally, we need the following notion of graph expansion, which was introduced by Komlós and Szemerédi [20].

5.1. Graph expansion. For $\varepsilon_1 > 0$ and $t > 0$, define

$$\varepsilon(x) = \varepsilon(x, \varepsilon_1, t) := \begin{cases} 0 & \text{if } x < t/5 \\ \frac{\varepsilon_1}{\log^2(15x/t)} & \text{if } x \geq t/5, \end{cases} \quad (5.1)$$

where, when it is clear from context we will not write the dependency on ε_1 and t of $\varepsilon(x)$. Note that, for $x \geq t/2$, $\varepsilon(x)$ is decreasing, while $\varepsilon(x) \cdot x$ is increasing.

Definition 5.1 ((ε_1, t) -expander). *A graph G is an (ε_1, t) -expander if all subsets $X \subseteq V(G)$ of size $t/2 \leq |X| \leq |G|/2$ satisfy*

$$|\Gamma_G(X)| \geq \varepsilon(|X|) \cdot |X|.$$

We will use the following lemma, essentially proved by Komlós and Szemerédi [20, 21], which states that every graph G contains an (ε_1, t) -expander subgraph H whose average degree is almost as large as that of G .

Lemma 5.2 ([20, 21]). *Let $C > 12, \varepsilon_1 \leq 1/(10C), c' < 1/2, d > 0$ and $\varepsilon(x) = \varepsilon(x, \varepsilon_1, c'd)$ as in (5.1). Then every graph G with $d(G) = d$ has a subgraph H such that*

- (i) H is an $(\varepsilon_1, c'd)$ -expander;
- (ii) $d(H) \geq (1 - \varepsilon_0)d$, where $\varepsilon_0 := \frac{C\varepsilon_1}{\log 3} < 1$;
- (iii) $\delta(H) \geq d(H)/2$;
- (iv) H is νd -connected, where $\nu := \frac{\varepsilon_1}{6 \log^2(5/c')}$.

Proof. Parts (i)–(iii) were shown in [21]. We only need to show H has high connectivity. Suppose H has a vertex cut S of size less than νd , where $\nu = \varepsilon_1/(6 \log^2(5/c'))$. Let X be the smallest component in $H - S$. Then $x := |X| < |H|/2$. On the other hand, for any vertex $v \in X$, we have $\Gamma_H(\{v\}) \subseteq X \cup S$. Since $\delta(H) \geq (1 - \varepsilon_0)d/2$, we have that

$$|X| \geq \delta(H) - |S| > \frac{(1 - \varepsilon_0)d}{2} - \nu d \geq \frac{d}{3} \geq \frac{c'd}{2}.$$

The expansion property (i) of H implies that $|\Gamma_H(X)| \geq \varepsilon(x)x$. Thus, since $\Gamma_H(X) \subseteq S$ and $\varepsilon(x)x$ is increasing for $x \geq c'd/2$, we have

$$|S| \geq |\Gamma_H(X)| \geq \varepsilon(x)x \geq \varepsilon\left(\frac{d}{3}\right) \frac{d}{3} = 2\nu d > 2|S|,$$

a contradiction. □

We remark that the expander subgraph H found in Lemma 5.2 could be much smaller than G , e.g. when G is a disjoint union of small cliques. The following property of expanders (Corollary 2.3 in [21]) is the only one which we require in our proof. It states that any two sets, provided that they are sufficiently large, are connected by a relatively short path, even after deleting an arbitrary small set of vertices.

Lemma 5.3 ([21]). *Let $\varepsilon_1, t > 0$ and let H be an n -vertex (ε_1, t) -expander and $X, X', W \subseteq V(H)$. If $|X|, |X'| \geq x \geq t/2$ and $|W| \leq \varepsilon(x)x/4$, then there is a path in $H - W$ from X to X' of length at most*

$$\frac{2}{\varepsilon_1} \log^3 \left(\frac{15n}{t} \right).$$

Throughout the rest of this section, we will set

$$t := c'd := \frac{d}{30} \quad \text{and write} \quad \varepsilon(x) := \varepsilon\left(x, \varepsilon_1, \frac{d}{30}\right).$$

(So $\varepsilon(x) = \varepsilon_1(\log^2(450x/d))^{-1}$ if $x \geq d/150$ and $\varepsilon(x) = 0$ otherwise.)

5.2. Hamiltonian subsets in expanders. The aim of this section and the next is to prove the following lemma, which states that expanders in which every vertex degree is not too far from d contain many Hamiltonian subsets.

Lemma 5.4. *Suppose that $0 < 1/d, 1/K \ll \varepsilon_1, 1/L \leq 1$ and $n \in \mathbb{N}$ is such that $n \geq Kd$. Let H be an n -vertex $(\varepsilon_1, d/30)$ -expander with $d/10 \leq \delta(H) \leq \Delta(H) \leq Ld$. Then $c(H) \geq 2^{50d}$.*

We split the proof of Lemma 5.4 into two cases depending on the edge-density of the expander graph; that is, how large d is compared to n . In the case when d is fairly large compared to n , we prove Lemma 5.4 in Section 5.3. The proof for the remaining case may be found in Section 5.4.

For the remainder of this section, our aim is to prove Lemma 5.7, the main ingredient for the first case. Roughly speaking, it states that almost regular expanders contain a large collection of ‘webs’ with certain properties. A web (see Definition 5.6) is a special tree which we will use in Section 5.3 to construct many Hamiltonian subsets in the case when d is not too small compared to n . A web contains many special subtrees which we call *units*, which themselves contain many large stars. By a k -star, we mean a star with k leaves.

Definition 5.5 ((h_1, h_2, h_3) -unit). *Given integers $h_1, h_2, h_3 > 0$, we define $F = F_u = \bigcup_{i \in [h_1]} P_i \cup \bigcup_{i \in [h_1]} S(x_i)$ to be an (h_1, h_2, h_3) -unit if it satisfies the following.*

- F contains distinct vertices u (the core vertex of F) and x_1, \dots, x_{h_1} .
- $\{P_i : i \in [h_1]\}$ is a collection of pairwise internally vertex-disjoint paths, each of length at most h_3 , such that P_i is a u, x_i -path.
- $\{S(x_i) : i \in [h_1]\}$ is a collection of vertex-disjoint h_2 -stars such that $S(x_i)$ has centre x_i and $\bigcup_{i \in [h_1]} (S(x_i) - \{x_i\})$ is vertex-disjoint from $\bigcup_{i \in [h_1]} P_i$.

Define the exterior $\text{Ext}(F) := \bigcup_{i \in [h_1]} (V(S(x_i)) - \{x_i\})$ and interior $\text{Int}(F) := V(F) \setminus \text{Ext}(F)$. For every vertex $w \in \text{Ext}(F)$, let $P(F, w)$ be the unique u, w -path in F .

A web is then defined via units as follows.

Definition 5.6 ((h_0, h_1, h_2, h_3) -web). *Given integers $h_0, h_1, h_2, h_3 > 0$, we define $W = \bigcup_{i \in [h_0]} Q_i \cup \bigcup_{i \in [h_0]} F_{u_i}$ to be an (h_0, h_1, h_2, h_3) -web if it satisfies the following.*

- W contains distinct vertices v (the core vertex of W) and u_1, \dots, u_{h_0} .
- $\{Q_i : i \in [h_0]\}$ is a collection of pairwise internally vertex-disjoint paths such that Q_i is a v, u_i -path of length at most h_3 .
- $\{F_{u_i} : i \in [h_0]\}$ is a collection of vertex-disjoint (h_1, h_2, h_3) -units such that F_{u_i} has core vertex u_i and $\bigcup_{i \in [h_0]} (F_{u_i} - \{u_i\})$ is vertex-disjoint from $\bigcup_{i \in [h_0]} Q_i$.

Define the exterior $\text{Ext}(W) := \bigcup_{i \in [h_0]} \text{Ext}(F_{u_i})$, interior $\text{Int}(W) := V(W) \setminus \text{Ext}(W)$ and centre $\text{Cen}(W) := \bigcup_{i \in [h_0]} V(Q_i)$. For every vertex $w \in \text{Ext}(W)$, let $P(W, w)$ be the unique v, w -path in W .

These structures are illustrated in Figure 3. Note that a web W with core vertex v is a tree with root v , and $\text{Cen}(W) \subset \text{Int}(W)$.

Throughout Sections 5.2 and 5.3, we will define a parameter m as follows and assume that:

$$m := \frac{2}{\varepsilon_1} \log^3 \left(\frac{450n}{d} \right); \quad 0 < \frac{1}{d}, \frac{1}{K} \ll \varepsilon_1, \frac{1}{L} \leq 1; \quad \text{and} \quad \log^{100} n \leq d \leq \frac{n}{K}.$$

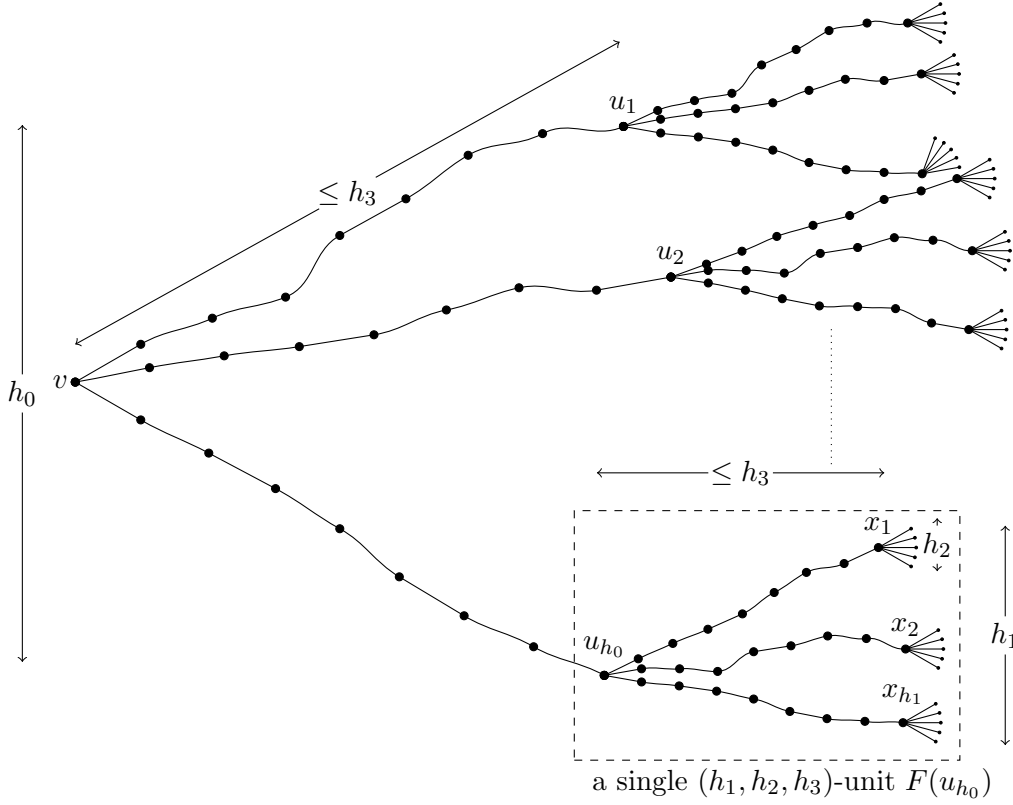
These assumptions imply the following useful inequalities (whose derivations we omit). If $d/30 \leq x \leq n$, then

$$\varepsilon(x) \geq \varepsilon(n) > \frac{1}{m}; \quad \text{and also} \quad n \geq Ldm^{100} \quad \text{and} \quad d \geq m^{30}. \quad (5.2)$$

The following lemma guarantees a large collection of webs with disjoint interiors and is a key ingredient in Section 5.3.

Lemma 5.7. *Suppose that $0 < 1/d, 1/K \ll \varepsilon_1, 1/L \leq 1$ and $n, d \in \mathbb{N}$ with $\log^{100} n \leq d \leq n/K$. Let H be an n -vertex $(\varepsilon_1, d/30)$ -expander with $d/10 \leq \delta(H) \leq \Delta(H) \leq Ld$. Then H contains $200d$ webs W_1, \dots, W_{200d} such that the following hold.*

- (i) W_i is an $(m^3, m^3, d/100, 4m)$ -web for all $i \in [200d]$;
- (ii) $\text{Int}(W_1), \dots, \text{Int}(W_{200d})$ are pairwise disjoint sets of vertices.

FIGURE 3. An (h_0, h_1, h_2, h_3) -web W .

One cannot hope that the webs themselves are disjoint. Indeed, the sum of the size of the exteriors of the desired collection of webs is at least $200d \cdot m^3 \cdot m^3 \cdot d/100 = 2d^2m^6$, which is much larger than n when $d > \sqrt{n}$.

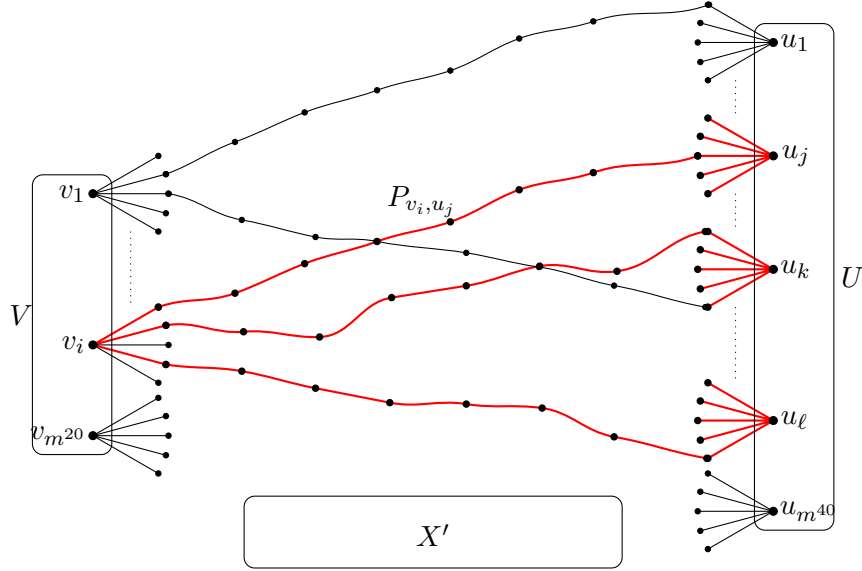
We will prove Lemma 5.7 in the following two subsections. First we use many disjoint stars to construct a large collection of disjoint units (Lemma 5.8). We then use these units to build the desired webs.

5.2.1. From stars to units.

Lemma 5.8. *Suppose that $0 < 1/d, 1/K \ll \varepsilon_1, 1/L \leq 1$ and $n, d \in \mathbb{N}$ with $\log^{100} n \leq d \leq n/K$. Let $h_1 \in [m^{10}]$ and $h_2 \in [d/100]$. Suppose that H is an n -vertex $(\varepsilon_1, d/30)$ -expander with $d/10 \leq \delta(H) \leq \Delta(H) \leq Ld$. Let $X \subseteq V(H)$ have size at most dm^{10} . Then $H - X$ contains a collection of $dm^{15}/(h_1h_2)$ pairwise vertex-disjoint $(h_1, h_2, m+2)$ -units.*

We briefly sketch the proof of Lemma 5.8. Suppose we have already found some vertex-disjoint units and wish to find another, F , to add to the collection. Remove those vertices in X' used in the units that we have already found. Our graph $H - X'$ contains a collection \mathcal{S} of many large disjoint stars since it is still almost regular. We partition the set of centres of stars in \mathcal{S} into two groups U and V . Then take a maximal collection \mathcal{P} of short paths in $H - X'$ which go between U and V , whose interiors are disjoint and do not intersect U or V . Note that these paths could intersect the leaves of stars in \mathcal{S} . We argue that this collection of paths can be extended unless there is some $v \in V$ which is the endpoint of every path in some large subset $\mathcal{P}' \subseteq \mathcal{P}$. Let $U' \subseteq U$ be the non- v endpoints of paths in \mathcal{P}' . Let v be the core vertex of F and the vertices I in paths of \mathcal{P}' form the interior of F . If a leaf of a U' -centred star lies in I , remove it. Few such leaves are removed since the paths in \mathcal{P}' are short and the stars of \mathcal{S} are large. The remaining leaves form the exterior of F . Figure 4 illustrates the proof of Lemma 5.8.

Proof of Lemma 5.8. Note that $dm^{15}/(h_1h_2)$ vertex-disjoint $(h_1, h_2, m+2)$ -units together contain at most $(h_1h_2 + (m+2)h_1)dm^{15}/(h_1h_2) \leq 1.1dm^{16}$ vertices. Thus it suffices to show that we can find an $(h_1, h_2, m+2)$ -unit in $H - X'$ for an arbitrary $X' \subseteq V(H)$ of size at most $2dm^{16}$ (then one

FIGURE 4. The proof of Lemma 5.8: a unit with core vertex $v_i \in V$ which avoids X' .

can take X' to be the union of X and at most $1.1dm^{16}$ vertices in the units we have already built). Fix such a set X' . We first find many vertex-disjoint stars in $H - X'$.

Claim 5.9. $H - X'$ contains a collection of m^{50} pairwise vertex-disjoint $d/50$ -stars.

Proof: Consider a maximal collection \mathcal{S} of vertex-disjoint $d/50$ -stars in $H - X'$ and let X'' be the set of vertices spanned by the stars of \mathcal{S} . If \mathcal{S} contains less than m^{50} stars, then $|X''| \leq dm^{50}$.

Then the fact that $d/10 \leq \delta(H) \leq \Delta(H) \leq Ld$ implies that

$$d(H - X' - X'') \geq \frac{n \cdot \delta(H) - |X' \cup X''| \cdot 2\Delta(H)}{n} \geq \frac{d}{10} - \frac{2dm^{50} \cdot 2Ld}{n} \stackrel{(5.2)}{\geq} \frac{d}{50}.$$

Thus $H - X' - X''$ contains a $d/50$ -star, contradicting to the maximality of \mathcal{S} . ■

By Claim 5.9, we can find $m^{20} + m^{40}$ pairwise vertex-disjoint $d/50$ -stars $S(v_1), \dots, S(v_{m^{20}})$ and $S(u_1), \dots, S(u_{m^{40}})$ in $H - X'$ such that $S(x)$ has centre x for all $x \in \{v_i, u_j : i \in [m^{20}], j \in [m^{40}]\}$. Set $V := \{v_1, \dots, v_{m^{20}}\}$ and $U := \{u_1, \dots, u_{m^{40}}\}$. Let \mathcal{P} be a maximal subset of $U \times V$ such that a collection $\mathcal{Q} := \{P_{v_i, u_j} : (v_i, u_j) \in \mathcal{P}\}$ of paths with the following properties exists.

- (C1) P_{v_i, u_j} is a v_i, u_j -path in H of length at most $m + 2$ which contains an edge of $S(u_j)$;
- (C2) for each $(v_i, u_j) \in \mathcal{P}$, $\text{Int}(P_{v_i, u_j})$ is disjoint from $X' \cup U \cup V$;
- (C3) all paths in \mathcal{Q} are pairwise internally vertex-disjoint.

Let $\Gamma_{\mathcal{P}}(v_i) := \{u_j : (v_i, u_j) \in \mathcal{P}\}$. We claim that there is a $v_i \in V$ such that

$$|\Gamma_{\mathcal{P}}(v_i)| \geq m^{10}. \quad (5.3)$$

Suppose to the contrary that $|\Gamma_{\mathcal{P}}(v_i)| < m^{10}$ for all $i \in [m^{20}]$. Let

$$\begin{aligned} P' &:= \bigcup_{(v_i, u_j) \in \mathcal{P}} \text{Int}(P_{v_i, u_j}); \quad J := \{j \in [m^{40}] : V(S(u_j)) \cap P' = \emptyset\}; \\ A &:= \bigcup_{i \in [m^{20}]} (V(S(v_i)) - \{v_i\}); \quad B := \bigcup_{j \in J} (V(S(u_j)) - \{u_j\}) \quad \text{and} \quad W := X' \cup U \cup V \cup P'. \end{aligned}$$

We will construct a path between A and B which avoids W by using Lemma 5.3 to contradict the maximality of \mathcal{P} . In order to do this, we estimate the sizes of A, B and W . We have

$$|P'| \leq \sum_{(v_i, u_j) \in \mathcal{P}} |\text{Int}(P_{v_i, u_j})| \stackrel{(C1)}{\leq} \sum_{(v_i, u_j) \in \mathcal{P}} m \leq m \sum_{i \in [m^{20}]} |\Gamma_{\mathcal{P}}(v_i)| \leq m \cdot m^{10} \cdot m^{20} = m^{31}. \quad (5.4)$$

This implies that

$$|J| \geq m^{40} - |P'| \geq m^{40} - m^{31} \geq \frac{m^{40}}{2}. \quad (5.5)$$

Since the $S(v_i)$ are vertex-disjoint $d/50$ -stars, we have

$$|A| = \left| \bigcup_{i \in [m^{20}]} (V(S(v_i)) - \{v_i\}) \right| \geq m^{20} \cdot \frac{d}{50} \geq dm^{19}, \quad (5.6)$$

Also (5.5) implies that

$$|B| = \left| \bigcup_{j \in J} (V(S(u_j)) - \{u_j\}) \right| \geq \frac{m^{40}}{2} \cdot \frac{d}{50} \geq dm^{19}. \quad (5.7)$$

Recall that, by (5.2), $d \geq m^{30}$. Thus

$$|W| = |X'| + |U| + |V| + |P'| \stackrel{(5.4)}{\leq} 2dm^{16} + m^{40} + m^{20} + m^{31} \stackrel{(5.2)}{\leq} dm^{17} < \frac{1}{4}\varepsilon(dm^{19})dm^{19}.$$

This together with (5.6) and (5.7) allows us to apply Lemma 5.3 with A, B, W and dm^{19} playing the roles of X, X', W and x respectively. Then we can find a minimal path Q of length at most $\frac{2}{\varepsilon_1} \log^3(450n/d) = m$ in $H - W$ from A to B . Then there exists $i' \in [m^{20}]$ and $j' \in [m^{40}]$ such that each endpoint of Q belongs to $S(v_{i'})$ and $S(u_{j'})$. Then $H[Q \cup \{v_{i'}, u_{j'}\}]$ contains a path $P_{v_{i'}, u_{j'}}$ from $v_{i'}$ to $u_{j'}$. Note that $P_{v_{i'}, u_{j'}}$ has length at most $m + 2$.

Note that $(v_{i'}, u_{j'}) \notin \mathcal{P}$ because $u_{j'} \in J$ and \mathcal{Q} satisfies (C1). Let $\mathcal{P}' := \mathcal{P} \cup \{(v_{i'}, u_{j'})\}$ and $\mathcal{Q}' := \mathcal{Q} \cup \{P_{v_{i'}, u_{j'}}\}$. We claim that $(\mathcal{P}', \mathcal{Q}')$ satisfies (C1)–(C3). Indeed, $(\mathcal{P}', \mathcal{Q}')$ satisfies (C1) since every path in \mathcal{Q} as well as $P_{v_{i'}, u_{j'}}$ has length at most $m + 2$ and contains an edge of $S(u_{j'})$. Also (C2) holds because $\text{Int}(P_{v_{i'}, u_{j'}}) \subseteq Q$ is disjoint from $W = X' \cup U \cup V \cup P'$ and \mathcal{Q} satisfies (C2). Finally (C3) is satisfied since $\text{Int}(P_{v_{i'}, u_{j'}}) \subseteq Q$ is disjoint from $P' = \bigcup_{(v_i, u_j) \in \mathcal{P}} \text{Int}(P_{v_i, u_j})$ and \mathcal{Q} satisfies (C3). This contradicts the maximality of \mathcal{P} , and thus there exists a vertex $v_i \in V$ satisfying (5.3).

Now we use v_i to construct an $(h_1, h_2, m + 2)$ -unit as desired (see Figure 4). Let $U' \subseteq N_{\mathcal{P}}(v_i)$ be a subset of U such that $|U'| = h_1 \leq m^{10}$. Let $u_j \in N_{\mathcal{P}}(v_i)$ be arbitrary. Then by (C2) $u_j \notin \text{Int}(P_{v_i, u_{j'}})$ for any $u_{j'} \in U'$, and also

$$\left| V(S(u_j)) \setminus \left(\bigcup_{u_{j'} \in U'} \text{Int}(P_{v_i, u_{j'}}) \right) \right| \geq \frac{d}{50} - m^{10} \cdot m \geq \frac{d}{100} \geq h_2.$$

Thus $V(S(u_j)) \setminus (\bigcup_{u_{j'} \in U'} \text{Int}(P_{v_i, u_{j'}}))$ contains an h_2 -star $S'(u_j)$ and $\bigcup_{u_j \in U'} (P_{v_i, u_j} \cup S'(u_j))$ forms an $(h_1, h_2, m + 2)$ -unit disjoint from X' . This finishes the proof. \square

5.2.2. From units to webs. We can now use Lemma 5.8 to prove Lemma 5.7 (and the proof itself is similar).

Proof of Lemma 5.7. We iteratively construct $(m^3, m^3, d/100, 4m)$ -webs W_1, \dots, W_{200d} such that $\text{Int}(W_i) \cap \text{Int}(W_j) = \emptyset$ for all $ij \in \binom{[200d]}{2}$. Suppose that we have constructed W_1, \dots, W_t for some $t < 200d$. Let $X := \bigcup_{i \in [t]} \text{Int}(W_i)$. Then

$$|X| \leq 200d \cdot m^3 \cdot m^3 \cdot 8m \leq dm^8. \quad (5.8)$$

Now we construct an $(m^3, m^3, d/100, 4m)$ -web W_{t+1} in $H - X$. Apply Lemma 5.8 with $H, X, m^5, d/100$ playing the roles of H, X, h_1, h_2 . Since $dm^{15}(m^5 \cdot d/100)^{-1} \geq m^{10} + m^5$, this implies that $H - X$ contains vertex-disjoint $(m^5, d/100, m + 2)$ -units $F_{v_1}, \dots, F_{v_{m^5}}$ and $F_{u_1}, \dots, F_{u_{m^{10}}}$ such that F_w has core vertex w for all $w \in \{v_i, u_j : i \in [m^5], j \in [m^{10}]\}$.

Set $V := \{v_1, \dots, v_{m^5}\}$ and $U := \{u_1, \dots, u_{m^{10}}\}$. Let $\mathcal{P} \subseteq V \times U$ be a maximal subset such that a collection $\mathcal{Q} := \{P_{v_i, u_j} : (v_i, u_j) \in \mathcal{P}\}$ of paths with the following properties exists. Recall the definition of $P(F, w)$ given in Definition 5.5.

- (D1) P_{v_i, u_j} is a v_i, u_j -path of length at most $4m$ such that $P_{v_i, u_j} = P(F_{v_i}, v_{ij}) \cup P_{ij}^* \cup P(F_{u_j}, u_{ij})$, where $v_{ij} \in \text{Ext}(F_{v_i})$, $u_{ij} \in \text{Ext}(F_{u_j})$ and P_{ij}^* is a v_{ij}, u_{ij} -path;

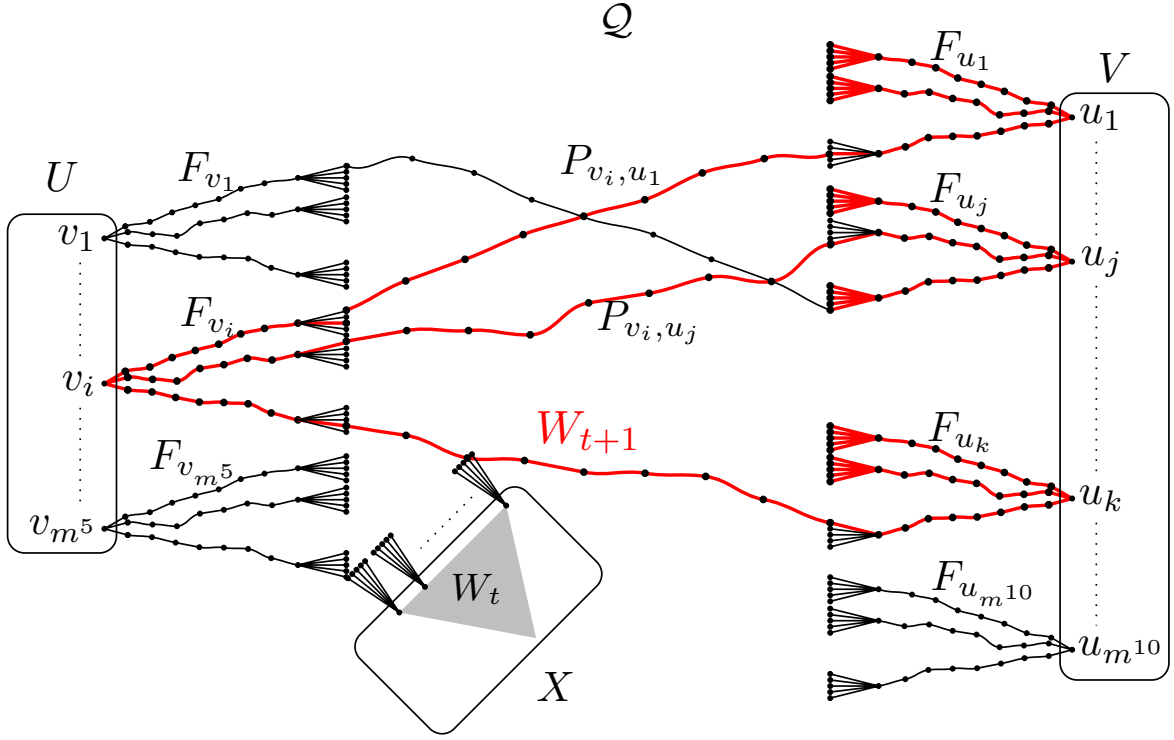


FIGURE 5. The proof of Lemma 5.7: a web W_{t+1} with core vertex v_i whose interior avoids X (the union of the interiors of W_1, \dots, W_t). Observe that W_{t+1} can and typically does intersect with the exteriors $\bigcup_{i \in [t]} \text{Ext}(W_i)$ of previously-built webs, but it does not intersect with the interiors $\bigcup_{i \in [t]} \text{Int}(W_i)$ of previously-built webs.

- (D2) for each $(v_i, u_j) \in \mathcal{P}$, P_{ij}^* is of length at most m and disjoint from $X \cup U \cup \bigcup_{k \in [m^5]} \text{Int}(F_{v_k})$, and P_{v_i, u_j} is disjoint from X ;
 (D3) all paths in \mathcal{Q} are pairwise internally vertex-disjoint.

Let $N_{\mathcal{P}}(v_i) := \{u_j : (v_i, u_j) \in \mathcal{P}\}$. We claim that there is a $v_i \in V$ such that

$$|N_{\mathcal{P}}(v_i)| \geq m^3. \quad (5.9)$$

Suppose to the contrary that $|N_{\mathcal{P}}(v_i)| < m^3$ for all $i \in [m^5]$. Let

$$\begin{aligned} P' &:= \bigcup_{(v_i, u_j) \in \mathcal{P}} \text{Int}(P_{v_i, u_j}); \quad J := \{j \in [m^{10}] : V(F_{u_j}) \cap P' = \emptyset\}; \\ A &:= \bigcup_{i \in [m^5]} \{w \in \text{Ext}(F_{v_i}) : V(P(F_{v_i}, w)) \cap V(P') = \emptyset\}; \quad B := \bigcup_{j \in J} \text{Ext}(F_{u_j}) \quad \text{and} \\ C &:= P' \cup X \cup U \cup \bigcup_{k \in [m^5]} \text{Int}(F_{v_k}). \end{aligned}$$

We will construct a path between A and B which avoids C by using Lemma 5.3 to contradict the maximality of \mathcal{P} . In order to do this, we estimate the sizes of A, B and C .

Since each $P_{v_i, u_j} \in \mathcal{Q}$ has length at most $4m$ and no $v_i \in V$ satisfies (5.9), we have

$$|P'| \leq \sum_{i \in [m^5]} \sum_{u_j \in N_{\mathcal{P}}(v_i)} |\text{Int}(P_{v_i, u_j})| < m^5 \cdot m^3 \cdot 4m = 4m^9. \quad (5.10)$$

Thus we have

$$|J| \geq m^{10} - |P'| \geq m^{10} - 4m^9 \geq \frac{m^{10}}{2}. \quad (5.11)$$

For any $i \in [m^5]$, recall that $\text{Int}(F_{v_i})$ consists of m^5 pairwise internally vertex-disjoint paths. By (D1), (D2) and the fact that $|N_{\mathcal{P}}(v_i)| < m^3$, the set P' contains at most m^3 of these paths. Thus

there are at least $m^5 - m^3 \geq m^5/2$ unused paths, implying that

$$|A| \geq \sum_{i \in [m^5]} \frac{m^5}{2} \cdot \frac{d}{100} - |P'| \stackrel{(5.10)}{\geq} \frac{dm^{10}}{200} - 4m^9 \geq 10dm^9.$$

Also

$$|B| = \left| \bigcup_{j \in J} \text{Ext}(F_{u_j}) \right| = |J| \cdot m^5 \cdot \frac{d}{100} \stackrel{(5.11)}{\geq} \frac{dm^{15}}{200} \geq 10dm^9.$$

Then, by (5.2), (5.8) and (5.10),

$$\begin{aligned} |C| &= |P'| + |X| + |U| + \left| \bigcup_{k \in [m^5]} \text{Int}(F_{v_k}) \right| \leq 4m^9 + dm^8 + m^{10} + \sum_{k \in [m^5]} m^5(m+2) \\ &\leq 2dm^8 \leq \frac{1}{4}\varepsilon(10dm^9) \cdot 10dm^9. \end{aligned}$$

Thus, we can apply Lemma 5.3 with $A, B, C, 10dm^9$ playing the roles of X, X', W, x respectively to find a minimal path Q of length at most m in $H - C$ from $v \in A$ to $u \in B$. Let $i' \in [m^5]$ and $j' \in [m^{10}]$ be the indices such that $v \in \text{Ext}(F_{v_{i'}})$ and $u \in \text{Ext}(F_{u_{j'}})$. Then $P_{v_{i'}, u_{j'}} := Q \cup P(F_{v_{i'}}, v) \cup P(F_{u_{j'}}, u)$ is a $v_{i'}, u_{j'}$ -path of length at most

$$m + |P(F_{v_{i'}}, v)| + |P(F_{u_{j'}}, u)| \leq m + 2(m+3) \leq 4m.$$

Note that by the definitions of J and B and the fact that \mathcal{Q} satisfies (D1), we have $(v_{i'}, u_{j'}) \notin \mathcal{P}$. Let

$$\mathcal{P}' := \mathcal{P} \cup \{(v_{i'}, u_{j'})\} \text{ and } \mathcal{Q}' := \mathcal{Q} \cup \{P_{v_{i'}, u_{j'}}\}.$$

We claim that $(\mathcal{P}', \mathcal{Q}')$ satisfies (D1)–(D3). It is easy to see that (D1) holds; while (D2) and (D3) follow from (D1), $Q \cap C = \emptyset$ and the fact that all units, in particular $F_{v_{i'}}$ and $F_{u_{j'}}$, are in $H - X$. This contradicts the maximality of \mathcal{P} . Thus there exists a vertex $v_i \in V$ satisfying (5.9).

We can now construct the desired web (see Figure 5). Let $U' \subseteq N_{\mathcal{P}}(v_i)$ be a subset of U such that $|U'| = m^3$. Let $P'' := \bigcup_{u_j \in U'} \text{Int}(P_{v_i, u_j})$. Then $|P''| \leq |U'| \cdot 4m = 4m^4$. Recall that P'' does not contain any vertex in U by (D2) and that F_{u_j} is an $(m^5, d/100, m+2)$ -unit for each $u_j \in U'$. Thus $F_{u_j} - P''$ contains an $(m^3, d/100, m+2)$ -unit F'_{u_j} because $m^5 - 4m^4 \geq m^3$. Then $\bigcup_{u_j \in U'} (P_{v_i, u_j} \cup F'_{u_j})$ is an $(m^3, m^3, d/100, 4m)$ -web with core vertex v_i in $H - X$. This finishes the proof. \square

5.3. Proof of Lemma 5.4 when $d \geq \log^{100} n$. Using Lemma 5.7, we are now able to prove Lemma 5.4 assuming in addition that $d \geq \log^{100} n$. This is achieved via Theorem 1.5, the general theorem about cycles in expander graphs stated in the introduction. We should note that the assumption $d \geq \log^{100} n$ is only used implicitly in the proof by applications of (5.2). Before presenting the proof, we give a brief sketch. The desired cycle is found via an iterative procedure. By Lemma 5.7 we can find a collection of webs W_1, \dots, W_{200d} whose interiors are disjoint and let Z be the set of core vertices of these webs and $U \in \binom{Z}{100d}$. The aim is to find u_1, \dots, u_k , $k \geq 98d$, in U such that, for each $i \in [k]$, there is a short u_i, u_{i+1} -path P_i which avoids Z and the centres of any web not centred at u_i or u_{i+1} , and $\text{Int}(P_1), \dots, \text{Int}(P_i)$ are pairwise disjoint. If we can find such a path P_i for all $i \in [k]$ (where $u_{k+1} := u_1$), then the concatenation of these paths is the desired cycle C_U . To achieve this, at the i^{th} step, we ensure that most webs, including the current web centred at u_i , have few of their interior vertices in the paths P_1, \dots, P_{i-1} which we have already created. Then the exteriors of our webs are still large enough to exclude previously-built paths to find the next path P_i between u_i and some u_{i+1} as required.

Proof of Theorem 1.5. For given ε_1, L we choose d_0, K_0 so that

$$0 < \frac{1}{d_0}, \frac{1}{K_0} \ll \varepsilon_1, \frac{1}{L} \leq 1.$$

As in the rest of the section, we let $m := \frac{2}{\varepsilon_1} \log^3(\frac{450n}{d})$, thus (5.2) holds. Since $d \geq d_0, K \geq K_0$, Lemma 5.7 implies that we can find in H a collection W_1, \dots, W_{200d} of $(m^3, m^3, \frac{d}{100}, 4m)$ -webs whose

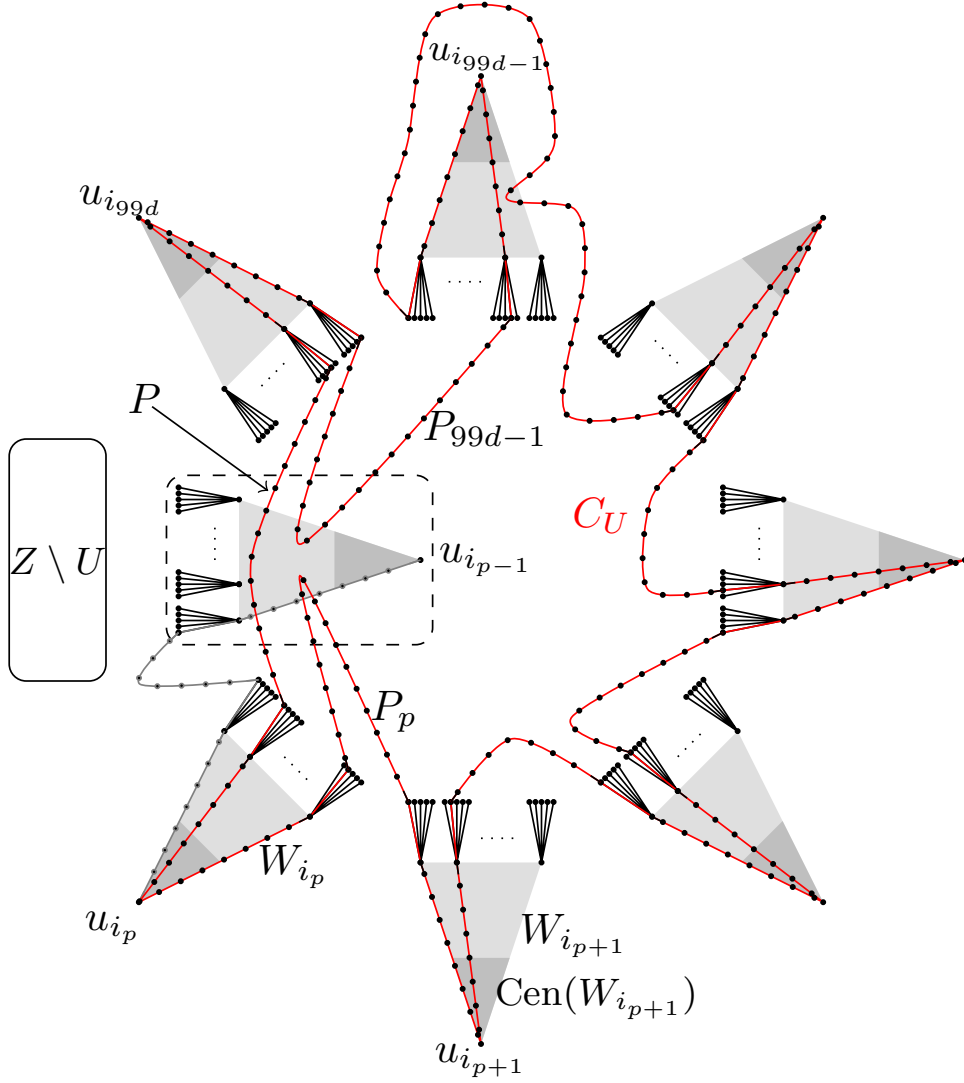


FIGURE 6. The proof of Theorem 1.5: constructing a cycle C_U . Here $p \in [99d]$ is the smallest index such that W_{i_p} is a good web (so the web $W_{i_{p-1}}$ enclosed in a dashed box is bad, i.e. its interior is over-used by the paths P_j). Paths P_j can intersect the interior of a web but not its centre.

interiors $\text{Int}(W_1), \dots, \text{Int}(W_{200d})$ are pairwise disjoint. Let $Z := \{u_1, \dots, u_{200d}\}$ where u_i is the core vertex of W_i for all $i \in [200d]$. Fix an arbitrary $100d$ -set U in Z . Without loss of generality, assume that $U = \{u_1, \dots, u_{100d}\}$. First, we show that there exists an index set $I = \{i_1, \dots, i_{99d}\} \subseteq [100d]$ and a collection $\mathcal{Q} = \{P_\ell : \ell \in [99d - 1]\}$ of paths satisfying the following. For each $\ell \in [99d - 1]$,

- (B1) P_ℓ is a $u_{i_\ell}, u_{i_{\ell+1}}$ -path of length at most $18m$;
- (B2) $\text{Int}(P_\ell)$ is disjoint from $\bigcup_{k \in [200d] \setminus \{i_\ell, i_{\ell+1}\}} \text{Cen}(W_k) \cup Z$;
- (B3) $\text{Int}(P_\ell)$ and $\text{Int}(P_k)$ are disjoint for all $k \in [99d - 1] \setminus \{\ell\}$;
- (B4) $|\text{Int}(W_{i_{\ell+1}}) \cap \bigcup_{k \in [\ell]} V(P_k)| < 2m^2$.

To find such an (I, \mathcal{Q}) , we will build a path between pairs in U avoiding vertices used in previously-built paths and the centres of all other webs. During the process, we will skip a web if its interior is ‘over-used’.

Assume we have built P_1, \dots, P_s and determined i_1, \dots, i_{s+1} with $s < 99d - 1$ satisfying (B1)–(B4). Since index set $\{1\}$ with the empty collection of paths satisfies (B1)–(B4), such a collection $\{P_1, \dots, P_s\}$ exists. Let $P' := \bigcup_{k \in [s]} \text{Int}(P_k)$. For $i \in [200d]$, we say a web W_i is *bad* if $|\text{Int}(W_i) \cap P'| \geq 2m^2$, and *good* otherwise. Note that (B4) implies that $W_{i_{s+1}}$ is good. By (B1),

$$|P'| \leq 18m \cdot s \leq 1800dm. \quad (5.12)$$

Claim 5.10. *Let W_{j_1} and W_{j_2} be two good webs. Then there exists a path P of length at most $18m$ in H from u_{j_1} to u_{j_2} such that $\text{Int}(P)$ is disjoint from $P' \cup Z \cup \bigcup_{k \in [200d] \setminus \{j_1, j_2\}} \text{Cen}(W_k)$.*

Proof: For $k \in [2]$, let $A_k := \{w \in \text{Ext}(W_{j_k}) : V(P(W_{j_k}, w)) \cap P' = \emptyset\}$. Since both W_{j_1}, W_{j_2} are good webs, P' intersects at most $2m^2$ of the paths of W_{j_k} in their interiors. Thus,

$$|A_k| \geq (m^3 - 2m^2) \cdot m^3 \cdot \frac{d}{100} - |P'| \stackrel{(5.12)}{\geq} \frac{dm^6}{200} - 1800dm \geq \frac{dm^6}{300}.$$

Note that $|\text{Cen}(W_k)| \leq 4m(m^3 + 1)$ for all $k \in [200d]$. Let

$$C := P' \cup \bigcup_{k \in [200d] \setminus \{j_1, j_2\}} \text{Cen}(W_k) \cup Z.$$

Thus (5.2) and (5.12) imply that

$$\begin{aligned} |C| &\leq |P'| + \sum_{k \in [200d] \setminus \{j_1, j_2\}} |\text{Cen}(W_k)| + |Z| \leq 1800dm + 200d \cdot 4m(m^3 + 1) + 200d \\ &\leq 1000dm^4 \leq \frac{1}{4} \cdot \varepsilon \left(\frac{dm^6}{300} \right) \cdot \frac{dm^6}{300}. \end{aligned}$$

Thus by applying Lemma 5.3 with $A_1, A_2, C, dm^6/300$ playing the roles of X, X', W, x respectively, there exists a path Q of length at most m in H between $w_1 \in A_1$ and $w_2 \in A_2$ avoiding C . Then $Q \cup P(W_{j_1}, w_1) \cup P(W_{j_2}, w_2)$ contains a u_{j_1}, u_{j_2} -path P and

$$|P| \leq |Q| + |P(W_{j_1}, w_1)| + |P(W_{j_2}, w_2)| \leq m + 2(4m + 4m + 1) \leq 18m.$$

By the choice of A_1, A_2 and the fact that the $\text{Cen}(W_k)$ sets are all pairwise disjoint, $\text{Int}(P)$ is disjoint from $P' \cup Z \cup \bigcup_{k \in [200d] \setminus \{j_1, j_2\}} \text{Cen}(W_k)$. \blacksquare

Since the interiors of W_1, \dots, W_{200d} are pairwise disjoint, (5.12) implies that the number of webs whose interiors contain at least m^2 vertices of P' is at most

$$\frac{1800dm}{m^2} < \frac{d}{2}. \quad (5.13)$$

Since $s < 99d - 1$, we can choose $i_{s+2} \in U \setminus \{i_1, \dots, i_{s+1}\}$ such that

$$|\text{Int}(W_{i_{s+2}}) \cap P'| \leq m^2. \quad (5.14)$$

Recall that $W_{i_{s+1}}$ is good. Thus by Claim 5.10, there is a $u_{i_{s+1}}, u_{i_{s+2}}$ -path P_{s+1} of length at most $18m$. Then it is easy to see that $\{i_1, \dots, i_{s+2}\}$ together with P_1, \dots, P_{s+1} satisfy (B1)–(B3) since $\text{Int}(P_{s+1})$ is disjoint from $P' \cup \bigcup_{k \in [200d] \setminus \{i_\ell, i_{\ell+1}\}} \text{Cen}(W_k) \cup Z$. Moreover,

$$\left| \text{Int}(W_{i_{s+2}}) \cap \bigcup_{k \in [s+1]} V(P_k) \right| = |\text{Int}(W_{i_{s+2}}) \cap P'| + |\text{Int}(W_{i_{s+2}}) \cap P_{s+1}| \stackrel{(5.14)}{\leq} m^2 + 18m < 2m^2,$$

so (B4) also holds. Therefore, we can repeat this process until $s = 99d - 1$, upon which we obtain the desired (I, \mathcal{Q}) satisfying (B1)–(B4).

Observe that, as before, (5.13) implies that less than $d/2$ indices $k \in [100d] \setminus I$ are such that W_k is bad. Let $p \in [99d]$ be the minimum index such that W_{i_p} is a good web (see Figure 6). Note that $W_{i_{99d}}$ is good by (B4). Then $p \leq d/2$ and so $|\{i_p, i_{p+1}, \dots, i_{99d}\}| > 98d$. By (B1), (B2) and (B3), the concatenation of $P_p, P_{p+1}, \dots, P_{99d-1}$ is a $u_{i_p}, u_{i_{99d}}$ -path avoiding $Z \setminus U$. By Claim 5.10, there exists a $u_{i_p}, u_{i_{99d}}$ -path P of length at most $18m$ such that $\text{Int}(P)$ is disjoint from $\bigcup_{k \in [99d-1]} \text{Int}(P_k) \cup Z \cup \bigcup_{k \in [200d] \setminus \{i_p, i_{99d}\}} \text{Cen}(W_k)$. Thus, the concatenation of $P_p, P_{p+1}, \dots, P_{99d-1}, P$ form a cycle C_U , as in Figure 6. Finally, by (B1), (B2) and Claim 5.10,

$$V(C_U) \cap Z = \{u_{i_p}, u_{i_{p+1}}, \dots, u_{i_{99d}}\} \in \binom{U}{\geq 98d},$$

completing the proof of the theorem. \square

The proof of Lemma 5.4 in the case when $d \geq \log^{100} n$ now follows easily since there are not many distinct U, U' such that $C_U \cap Z$ and $C_{U'} \cap Z$ are equal. That is, the cycles C_U are ‘almost distinguishable’ by their intersection with Z .

Proof of Theorem 1.5 \Rightarrow Lemma 5.4 when $d \geq \log^{100} n$. Apply Theorem 1.5 to obtain a set $Z \subseteq V(H)$ of size $200d$ such that, for every subset $U \subseteq Z$ of size $100d$, there exists a cycle C_U with $V(C_U) \cap Z \in \binom{U}{\geq 98d}$.

Fix an arbitrary cycle C in H such that $V(C) \cap Z$ has size $98d \leq |V(C) \cap Z| \leq 100d$. There are at most $\binom{|Z| - |V(C) \cap Z|}{100d - |V(C) \cap Z|}$ ways to choose a $100d$ -set $U \subseteq Z$ containing $V(C) \cap Z$. In other words, for a fixed cycle C in H ,

$$|\{U \subseteq Z : V(C_U) = V(C)\}| \leq \binom{|Z| - |V(C) \cap Z|}{100d - |V(C) \cap Z|} \leq \binom{102d}{2d}.$$

Therefore the number of Hamiltonian subsets in H is

$$c(H) \geq \frac{\binom{|Z|}{100d}}{\binom{102d}{2d}} \geq 2^{50d},$$

as desired. \square

5.4. Proof of Lemma 5.4 when $d \leq \log^{100} n$. In this section, we finish the proof of Lemma 5.4 by proving it under the additional assumption that $d \leq \log^{100} n$. The proof will follow easily from the next result.

Lemma 5.11. *Let $0 < 1/d \ll \varepsilon_1, 1/L \leq 1$. Suppose that H is an n -vertex $(\varepsilon_1, d/30)$ -expander with $d \leq \log^{100} n$ and $d/10 \leq \delta(H) \leq \Delta(H) \leq Ld$. Then H contains a set Z of size $200d$ such that for every subset $U \subseteq Z$ of size $100d$, there exists a cycle C_U with $V(C_U) \cap Z = U$.*

First let us see how this implies Lemma 5.4. Let Z be the set guaranteed in Lemma 5.11. Then for distinct $100d$ -sets $U, U' \subseteq Z$, their corresponding cycles C_U and $C_{U'}$ can be distinguished in Z , implying that $c(H) \geq \binom{200d}{100d} > 2^{50d}$ as desired.

The idea of the proof of Lemma 5.11 is as follows. Since H has small maximum degree, we can choose Z to be a set of $200d$ vertices which are very far apart in H . Let $U := \{u_1, \dots, u_{100d}\} \subseteq Z$. We will obtain the desired cycle C_U via an iterative procedure. For each $i \in [100d]$ we will obtain a path P_i between u_i and u_{i+1} (where $u_{100d+1} := u_1$) which avoids the rest of Z and the interiors of previously-built paths. The concatenation of these paths is the desired cycle C_U . To enable the iteration to be completed, we ensure that there is a large set of vertices within a short distance of u_i , and similarly for u_{i+1} , which have not been used in previously-built paths. Lemma 5.3 implies that there is a short path P_i between these two large sets which avoids existing paths. We then extend P_i into a u_i, u_{i+1} -path, as required.

To prove Lemma 5.11, we need the following result from [27] (Lemma 5.5 when taking $s = 5$ and $q = 1$). To state it we need some notation. Given an integer $r \geq 0$, graph G and $X \subseteq V(G)$, define $B_G^r(X)$ to be the *ball of radius r around X* ; that is, the set of vertices at distance at most r from X in G . For $r \geq 1$, define $\Gamma_G^r(X)$ by setting

$$\Gamma_G^r(X) := B_G^r(X) \setminus B_G^{r-1}(X).$$

So $\Gamma_G^r(X)$ is the r^{th} sphere around X , i.e. the set of vertices at distance exactly r from X . If $v \in V(G)$, we write $\Gamma_G^r(v)$ as shorthand for $\Gamma_G^r(\{v\})$ and similarly abbreviate $B_G^r(\{v\})$.

Lemma 5.12 ([27]). *Let $0 < 1/d \ll \varepsilon_1 \leq 1$ and $n \in \mathbb{N}$ be such that $d \leq \log^{100} n$, and set $r := (\log \log n)^5$. Let H be an n -vertex $(\varepsilon_1, d/30)$ -expander with $\delta(H) \geq d/10$. Let $v \in V(H)$ and $u \in B_H^r(v)$, and let P be a shortest u, v -path in $B_H^r(v)$. Then $|B_{H'}^r(v)| \geq d \log^7 n$, where $H' := H - (V(P) \setminus \{v\})$.*

Proof of Lemma 5.11. Since $1/d \ll 1$ and $d \leq \log^{100} n$, we have $1/n \ll 1$. Let

$$r := (\log \log n)^5 \quad \text{and} \quad k := (\log n)^{7/8}.$$

Note that

$$\Delta(H) \leq Ld \leq d^2 \leq \log^{200} n. \tag{5.15}$$

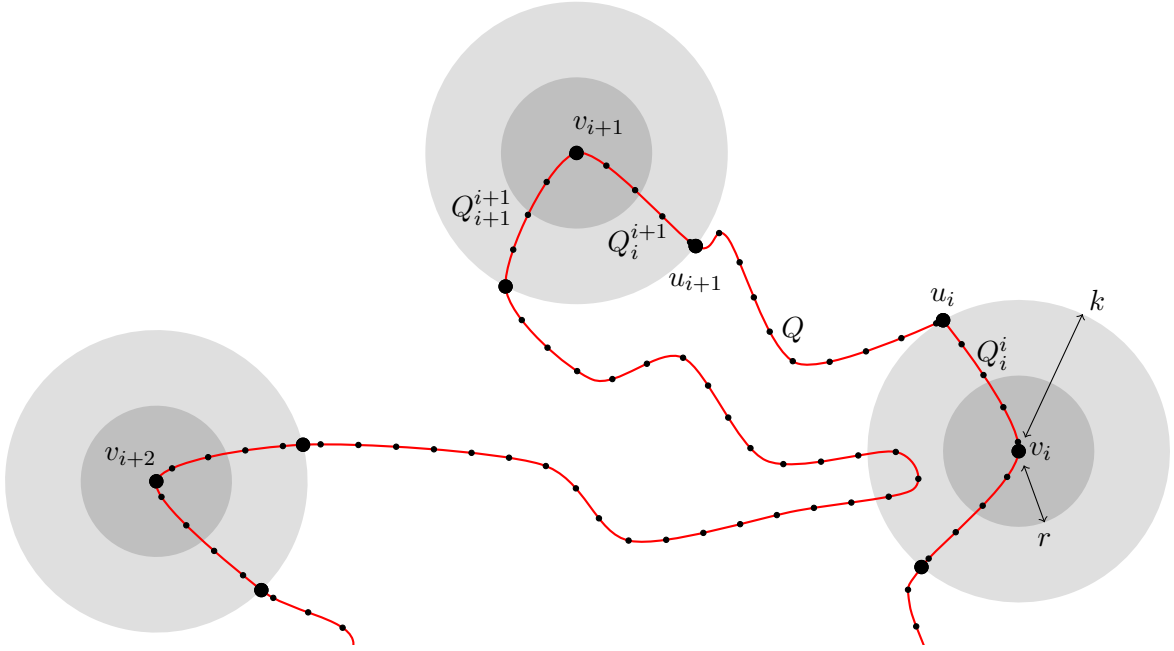


FIGURE 7. The proof of Lemma 5.11: the construction of $P_i = Q_i^i \cup Q \cup Q_i^{i+1}$. The red line is a segment of C_U , the concatenation of P_1, \dots, P_{100d} .

We first show that there exists a set of $200d$ vertices Z whose pairwise distances are all greater than $2k$. In other words, for all distinct $z, z' \in Z$, we have that $B_H^k(z) \cap B_H^k(z') = \emptyset$. Indeed, given such a set Z with $|Z| < 200d$, by (5.15) and the fact that $1/n \ll 1$,

$$|B^{2k}(Z)| \leq |Z| \cdot \sum_{0 \leq i \leq 2k} \Delta(H)^i < 2|Z| \cdot \Delta(H)^{2k} < 400d \cdot (\log^{200} n)^{2k} < n.$$

So we can add another vertex in $V(G) \setminus B^{2k}(Z)$ that is at distance greater than $2k$ from Z . Thus such a set Z with $|Z| = 200d$ exists.

By applying Lemma 5.12 with the empty path from v to v playing the role of P , for every $z \in Z$ we have that

$$|B_H^r(z)| \geq d \log^7 n. \quad (5.16)$$

Fix an arbitrary $100d$ -set $U \subseteq Z$, say $U = \{v_1, \dots, v_{100d}\}$, and let $v_{100d+1} := v_1$ and $v_0 := v_{100d}$. Let $I \subseteq [100d]$ be a maximal subset such that there exists a collection $\mathcal{Q} = \{P_j : j \in I\}$ satisfying the following properties.

- (E1) For each $j \in I$, P_j is a v_j, v_{j+1} -path with length at most $2 \log^4 n$;
- (E2) the paths in \mathcal{Q} are pairwise internally vertex-disjoint;
- (E3) for every $j \in I$, the path P_j is disjoint from $\bigcup_{p \in [100d] \setminus \{j, j+1\}} B_H^r(v_p) \cup (Z \setminus U)$;
- (E4) every $j \in I$ such that $j+1 \notin I$ satisfies $|B_{H-(V(P_j) \setminus \{v_{j+1}\})}^r(v_{j+1})| \geq d \log^7 n$;
- (E5) every $j \in I$ such that $j-1 \notin I$ satisfies $|B_{H-(V(P_j) \setminus \{v_j\})}^r(v_j)| \geq d \log^7 n$.

Note that such a maximal subset exists as by (5.16), we have that $(I, \mathcal{Q}) = (\{1\}, \emptyset)$ satisfies (E1)–(E5). We show next that $I = [100d]$. Indeed, suppose that there exists $i \in [100d] \setminus I$. Let

$$P' := \bigcup_{j \in I} \text{Int}(P_j), \quad W' := \bigcup_{p \in [100d] \setminus \{i, i+1\}} B_H^r(v_p), \quad \text{and} \quad W := P' \cup W' \cup (Z \setminus U).$$

We would like to estimate the sizes of P', W' and W . Firstly,

$$|P'| \stackrel{(E1)}{\leq} |I| \cdot 2 \log^4 n < 100d \cdot 2 \log^4 n \leq d \log^5 n. \quad (5.17)$$

By (5.15) and since $1/n \ll 1$,

$$|W'| \leq 100d \cdot \sum_{0 \leq \ell \leq r} (\log^{200} n)^\ell \leq 100 \log^{100} n \cdot 2 \log^{200r} n \leq \log^{300r} n = e^{300(\log \log n)^6} \leq e^{\log^{\frac{1}{5}} n}.$$

This implies that

$$|W| \leq |P'| + |W'| + |Z \setminus U| \leq d \log^5 n + e^{\log^{\frac{1}{5}} n} + 100d < \frac{1}{4} \varepsilon (e^{\log^{\frac{1}{4}} n}) e^{\log^{\frac{1}{4}} n}. \quad (5.18)$$

Note that (E3) implies that

$$\begin{aligned} B_{H-P'}^r(v_i) &= \begin{cases} B_H^r(v_i) & \text{if } i-1 \notin I \\ B_{H-(V(P_{i-1}) \setminus \{v_i\})}^r(v_i) & \text{if } i-1 \in I \end{cases} \quad \text{and} \\ B_{H-P'}^r(v_{i+1}) &= \begin{cases} B_H^r(v_{i+1}) & \text{if } i+1 \notin I \\ B_{H-(V(P_{i+1}) \setminus \{v_{i+1}\})}^r(v_{i+1}) & \text{if } i+1 \in I. \end{cases} \end{aligned} \quad (5.19)$$

Now we prove the following claim.

Claim 5.13. *For $p \in \{i, i+1\}$, we have that*

- (i) $|B_{H-P'}^r(v_p)| \geq d \log^7 n$;
- (ii) $|B_{H-P'}^k(v_p)| \geq e^{\log^{1/4} n}$.

Proof: Assertion (i) is a simple consequence of (5.16), (E4) and (E5). Indeed, (E4) and (E5) imply that, if $i-1 \in I$, then $|B_{H-P'}^r(v_i)| \geq d \log^7 n$; and if $i+1 \in I$, then $|B_{H-P'}^r(v_{i+1})| \geq d \log^7 n$. If $i-1 \notin I$, then $|B_{H-P'}^r(v_i)| = |B_H^r(v_i)| \geq d \log^7 n$ by (5.16) and (5.19), and similarly if $i+1 \notin I$, then $|B_{H-P'}^r(v_p)| \geq d \log^7 n$. This proves (i).

For (ii), let X be a set with $d \log^7 n \leq |X| \leq e^{\log^{1/4} n}$. Then, since $\varepsilon(x)$ is decreasing,

$$\varepsilon(|X|) \geq \varepsilon(e^{\log^{1/4} n}) = \frac{\varepsilon_1}{\log^2 \frac{450e^{\log^{1/4} n}}{d}} > \frac{\varepsilon_1}{\log^{1/2} n}. \quad (5.20)$$

Since H is an $(\varepsilon_1, d/30)$ -expander and $|X| \geq d \log^7 n$, if $X \cap P' = \emptyset$, then

$$|\Gamma_{H-P'}(X)| \geq |\Gamma_H(X)| - |P'| \geq \varepsilon(|X|)|X| - |P'| \stackrel{(5.17), (5.20)}{\geq} \frac{\varepsilon_1}{2 \log^{1/2} n} |X|. \quad (5.21)$$

Note that for $\ell \geq r$, (i) implies that $|B_{H-P'}^\ell(v_p)| \geq |B_{H-P'}^r(v_p)| \geq d \log^7 n$. Thus, if $|B_{H-P'}^\ell(v_p)| < e^{\log^{1/4} n}$, then, as $B_{H-P'}^\ell(v_p) \cap P' = \emptyset$, applying (5.21) with $X = B_{H-P'}^\ell(v_p)$ we have

$$|B_{H-P'}^{\ell+1}(v_p)| = |B_{H-P'}^\ell(v_p)| + |\Gamma_{H-P'}(B_{H-P'}^\ell(v_p))| \stackrel{(5.21)}{\geq} \left(1 + \frac{\varepsilon_1}{2 \log^{1/2} n}\right) |B_{H-P'}^\ell(v_p)|.$$

Thus, if $|B_{H-P'}^\ell(v_p)| < e^{\log^{1/4} n}$ for all $\ell < k$, then

$$\log |B_{H-P'}^k(v_p)| \geq \log \left(1 + \frac{\varepsilon_1}{2 \log^{1/2} n}\right)^{k-r} + \log |B_{H-P'}^r(v_p)| \geq \frac{\varepsilon_1(k-r)}{4 \log^{1/2} n} > \log^{1/4} n,$$

completing the proof of (ii). ■

Observe that, for $p \in \{i, i+1\}$, as all the balls $B_H^k(z)$ with $z \in Z$ are pairwise disjoint, the definitions of W' and W imply that

$$|B_{H-W}^k(v_p)| = |B_{H-P'}^k(v_p)| \geq e^{\log^{1/4} n}. \quad (5.22)$$

Together with (5.18), this allows us to apply Lemma 5.3 with $B_{H-W}^k(v_i), B_{H-W}^k(v_{i+1}), W$ and $e^{\log^{1/4} n}$ playing the roles of X, X', W and x respectively to show the existence of a path of length $\log^4 n$ between $B_{H-W}^k(v_i)$ and $B_{H-W}^k(v_{i+1})$ in $H - W$. Let Q be a shortest such path, $u_p := B_{H-W}^k(v_p) \cap Q$ and Q_i^p be a shortest path from v_p to u_p in $H - W$ for $p \in \{i, i+1\}$ (see Figure 7). Now define

$$P_i := Q_i^i \cup Q \cup Q_i^{i+1}, \quad I' := I \cup \{i\} \quad \text{and} \quad \mathcal{Q}' := \mathcal{Q} \cup \{P_i\}.$$

Note that $|P_i| \leq |Q| + 2k \leq 2 \log^4 n$, thus \mathcal{Q}' satisfies (E1). Since $P_i \cap W = \emptyset$, \mathcal{Q}' satisfies (E2) and (E3). For (E4), set $P'' = P' \cup (P_i \setminus \{v_{i+1}\})$. If $i+1 \notin I'$, then $i+1 \notin I$, so (5.19) and (5.22) imply that $B_{H-W}^r(v_{i+1}) = B_{H-P'}^r(v_{i+1}) = B_H^r(v_{i+1})$. This, together with the fact that $Q_i^i \cup Q$ is disjoint from $B_{H-W}^r(v_{i+1})$, implies that

$$B_{H-P''}^r(v_{i+1}) = B_{H-(V(P_i) \setminus \{v_{i+1}\})}^r(v_{i+1}) = B_{H-(V(Q_i^{i+1}) \setminus \{v_{i+1}\})}^r(v_{i+1}).$$

Since Q_i^{i+1} is a shortest path between v_{i+1} and u_{i+1} in $H - W$, it is also shortest in H by the definition of W . Thus Lemma 5.12 with v_{i+1}, Q_i^{i+1} playing the roles of v and P implies that (E4) for $j = i$. For $j \in I' \setminus \{i\}$, (E4) holds since \mathcal{Q} satisfies (E4) and \mathcal{Q}' satisfies (E3). Similarly, (E5) is also satisfied. This contradicts the maximality of I . Thus $I = [100d]$.

We claim that the concatenation of P_1, \dots, P_{100d} forms the desired cycle C_U . That C_U is a cycle follows from (E1) and (E2). Furthermore, C_U contains no vertices in $Z \setminus U$ by (E3), and $U \subseteq V(C_U)$ by (E1). Thus $V(C_U) \cap Z = U$, as required. This completes the proof of the lemma, and hence the proof of Lemma 5.4. \square

6. THE PROOF OF THEOREM 1.3

We will need the following definitions.

Definition 6.1 (cut vertex, block, block graph, leaf block). A *cut vertex* x of a graph G is such that $G - x$ has more components than G . A *block* of G is a maximal subgraph $H \subseteq G$ such that no vertex of H is a cut vertex of H . The blocks of G form a partition of $E(G)$, and any two blocks have at most one vertex in common. Note that a block is an induced subgraph. The *block graph* $BL(G)$ is the graph whose vertex set is the set of blocks of G , where two blocks are joined by an edge in $BL(G)$ whenever they share a vertex in G . A *leaf block* of G is a block which contains exactly one cut vertex of G (so it has degree one in $BL(G)$). Since $BL(G)$ is always a forest, there are at least two leaf blocks of G unless G is 2-connected. Note also that every cut vertex has degree at least two in every block containing it, when that block is not a single edge.

Now we start the main proof.

Proof of Theorem 1.3. Let $0 < \alpha \leq 1$. We choose ε_1, K, L so that

$$0 < \frac{1}{d_0} \ll \frac{1}{K} \ll \frac{1}{L} \ll \varepsilon_1, \alpha \leq 1 \quad \text{where} \quad \varepsilon_1 \leq \frac{1}{130}. \quad (6.1)$$

Now, let $d \geq d_0$ be an integer. Assume that G is a counterexample to Theorem 1.3 such that $n := |G|$ is minimal. Then $d(G) \geq d$; $G \notin \{K_{d+1}, K_{d+1} * K_d\}$ and $c(G) < (2 - \alpha)2^{d+1}$. Since K_{d+1} is the only graph with average degree at least d on $d + 1$ vertices, we have $n \geq d + 2$.

Claim 6.2. *For any $V' \subsetneq V(G)$, we have $d(G[V']) \leq d$. Also we have $\delta(G) \geq d/2$.*

Proof: Suppose that there is some $V' \subsetneq V(G)$ for which $d(G[V']) > d$. Then $G[V'] \notin \{K_{d+1}, K_{d+1} * K_d\}$ since both K_{d+1} and $K_{d+1} * K_d$ have average degree exactly d . Thus the minimality of G implies that $G[V']$ is not a counterexample to Theorem 1.3. Then $c(G) \geq c(G[V']) \geq (2 - \alpha)2^{d+1}$, a contradiction.

Assume now that there exists a vertex $x \in V(G)$ such that $d_G(x) < d/2$. Then

$$d(G - x) = \frac{2e(G) - 2d_G(x)}{n - 1} > \frac{dn - d}{n - 1} = d,$$

a contradiction to the first part of the claim. \blacksquare

Next, we show that G contains a large 2-connected subgraph which essentially inherits the average degree of G . To see this, we need the following claim.

Claim 6.3. *If G is not 2-connected, then for any leaf block $F \subseteq G$, the following hold.*

- (i) $\delta_2(F) \geq d/2$; $|F| \geq d$ and $d(F) \geq d - 1$;
- (ii) If in addition $d + 1 \leq |F| \leq 1.19d$, then $c(F) \geq (1 - \alpha/2)2^{d+1}$.

Proof: Fix a leaf block $F \subsetneq G$ and let x be the corresponding cut vertex, i.e. the only vertex in F which has neighbours outside of F . Note that $\delta_2(F) \geq \delta(G) \geq d/2$ by Claim 6.2.

Let $G' := G - (V(F) \setminus \{x\})$. Let $n' := |G'|$. Then we have

$$n = |F| + n' - 1, \quad \text{and} \quad e(G) = e(F) + e(G').$$

By Claim 6.2, we have $d(F) \leq d$ and $d(G') \leq d$. Thus,

$$e(F) = e(G) - e(G') \geq dn/2 - dn'/2 = d(|F| - 1)/2.$$

Since every vertex of F has at most $|F| - 1$ neighbours in F , this implies that

$$|F| - 1 \geq d(F) \geq \frac{d(|F| - 1)}{|F|} = d - \frac{d}{|F|}.$$

Rearranging this, we see that

$$|F| \geq d, \quad \text{and} \quad d(F) \geq d - 1, \quad (6.2)$$

proving (i).

To prove (ii), suppose now that $d + 1 \leq |F| \leq 1.19d$. Since x is a cut vertex, we have

$$\delta(F) \geq \min\{\delta(G), d_F(x)\} \quad \text{and} \quad \delta(F - x) \geq \delta(G) - 1 \geq (d - 2)/2 \quad (6.3)$$

by Claim 6.2. If $d_F(x) \geq (d - 1)/2$, then (6.3) implies that $\delta(F) \geq (d - 1)/2$. Since $1/(d - 1) \ll \alpha$ and $|F| \leq 1.19d \leq 1.2(d - 1)$, we can apply Lemma 4.1 to F with $d - 1$ playing the role of d to see that $c(F) \geq (1 - \alpha/2)2^{|F|} \geq (1 - \alpha/2)2^{d+1}$ as desired.

So we may assume that $d_F(x) < (d - 1)/2$. We claim that $|F| \geq d + 2$. Indeed, if this is not so, then

$$(d - 1)|F| \stackrel{(6.2)}{\leq} 2e(F) \leq 2 \binom{|F| - 1}{2} + 2d_F(x) < (|F| - 1)(|F| - 2) + (d - 1) \leq (d - 1)|F|,$$

a contradiction. Thus, $|F - x| \geq d + 1$. Note that

$$d(F - x) = \frac{2e(F) - 2d_F(x)}{|F| - 1} \stackrel{(6.2)}{>} \frac{(d - 1)|F| - (d - 1)}{|F| - 1} = d - 1.$$

Then by (6.3) and the fact that $1/(d - 2) \ll \alpha$ and $|F| - 1 \leq 1.19d - 1 \leq 1.2(d - 2)$, we can apply Lemma 4.1 to $F - x$ with $d - 2$ playing the role of d to get $c(F) \geq c(F - x) \geq (1 - \alpha/2)2^{d+1}$. ■

The next claim guarantees a large 2-connected subgraph G_1 of G , which still has average degree at least roughly d .

Claim 6.4. *There exists $G_1 \subseteq G$ such that*

- (i) G_1 is 2-connected;
- (ii) $d - 1 \leq d(G_1) \leq 4d$;
- (iii) $\delta_2(G_1) \geq d/2$;
- (iv) $|G_1| \geq K^2 d$.

Proof: Suppose first that G is 2-connected. Then we set $G_1 := G$. Then $d(G_1) \leq 4d$, since otherwise Theorem 1.2 implies $c(G) \geq 2^{2d}$, a contradiction. By Claim 6.2, $\delta(G_1) \geq d/2$. We are left to show that $|G_1| = n \geq K^2 d$. Recall that $n \geq d + 2$. If $n \leq 1.2d$, then by Lemma 4.1, $c(G) \geq (1 - \alpha/2)2^n \geq (2 - \alpha)2^{d+1}$, a contradiction. Thus we can assume that $1.2d \leq n < K^2 d$. Applying Lemma 4.2 with G, d and K^2 playing the roles of G, d and K respectively, we have that $c(G) \geq 2^{(1+1/200)d} \geq (2 - \alpha)2^{d+1}$, a contradiction.

We may therefore assume that G is not 2-connected. Then G contains at least two leaf blocks G_1, G_2 , where $|G_1| \geq |G_2|$. Then by Claims 6.2 and 6.3(i), for $i \in [2]$ we have that

$$\delta_2(G_i) \geq d/2; \quad |G_i| \geq d \quad \text{and} \quad d - 1 \leq d(G_i) \leq d. \quad (6.4)$$

We may assume that

$$d \leq |G_2| \leq |G_1| < K^2 d,$$

otherwise G_1 is the desired subgraph, as above. We distinguish the following three cases.

Case 1: $|G_1| \geq 1.19d$.

Then applying Lemma 4.2 to G_1 with $d - 1$ and K^2 playing the roles of d and K respectively, we get $c(G_1) \geq 2^{(1+1/200)(d-1)} \geq (2 - \alpha)2^{d+1}$, a contradiction.

Case 2: $d + 1 \leq |G_2| \leq |G_1| \leq 1.19d$.

Then Claim 6.3(ii) implies that for $i \in [2]$, $c(G_i) \geq (1 - \alpha/2)2^{d+1}$. Thus $c(G) \geq c(G_1) + c(G_2) \geq (2 - \alpha)2^{d+1}$, a contradiction.

Case 3: $|G_2| = d$.

Note that since $d(G_2) \geq d-1$ from (6.4), G_2 is isomorphic to K_d . Let x_2 be the cut vertex of G in G_2 . Consider the graph $G^* := G - (V(G_2) \setminus \{x_2\})$. Then

$$d(G^*) = \frac{2e(G^*)}{|G^*|} = \frac{2e(G) - 2e(G_2)}{n - d + 1} \geq \frac{dn - d(d-1)}{n - d + 1} = d.$$

Since $c(G^*) \leq c(G) < (2-\alpha)2^{d+1}$, the minimality of G implies that G^* is isomorphic to either K_{d+1} or $K_{d+1} * K_d$. If $G^* \cong K_{d+1}$, then $G = K_{d+1} * K_d$, a contradiction. If $G^* \cong K_{d+1} * K_d$, then

$$c(G) = c(G_2) + c(G^*) = 2c(K_d) + c(K_{d+1}) \geq (2-\alpha)2^{d+1},$$

a contradiction. ■

Let $G_1 \subseteq G$ be the subgraph guaranteed by Claim 6.4. The following claim states that, if H is any subgraph of G_1 with large average degree, then the average degree remains large after removing any small subset of vertices U . The claim will then be used to find expander subgraphs of G_1 which are almost regular.

Claim 6.5. *Suppose that U is a subset of $V(G_1)$ such that $|U| \leq 10|G_1|/L$. Then $d(G_1 - U) \geq 0.9(d-1)$.*

Proof: We may assume that $|U| = 10|G_1|/L$. (Indeed, if not, we can take $10|G_1|/L - |U|$ vertices in $V(G_1) \setminus U$ of largest degree and add them to U to obtain U' such that $d(G_1 - U) \geq d(G_1 - U')$.) Let $B := V(G_1) \setminus U$ and suppose to the contrary that $d(G_1[B]) < 0.9(d-1)$. Claim 6.2 implies that $d(G_1[U]) \leq d$. Let $H := G_1[B, U]$. Then, since $|U| = 10|G_1|/L$ and $1/L \ll 1$,

$$\begin{aligned} 2e(H) &= 2(e(G_1) - e(G_1[U]) - e(G_1[B])) \\ &\geq (d-1)|G_1| - d|U| - 0.9(d-1)(|G_1| - |U|) \\ &\geq \frac{(d-1)|G_1|}{10} - \frac{10d|G_1|}{L} \geq \frac{d|G_1|}{20}. \end{aligned} \tag{6.5}$$

Thus $d(H) \geq d/20$. Let $t := |G_1|/|U| = L/10$.

If H contains a subgraph H' with $d(H') \geq 3d$, then Theorem 1.2 implies that

$$c(G) \geq c(H') \geq 2^{3d/2},$$

a contradiction. Thus we may thus assume that every subgraph of H has average degree at most $3d$. Let

$$B' := \left\{ v \in B : d_H(v) \geq \frac{d}{100} \right\}.$$

Then by the definition of B' , we have

$$e(G_1[U, B']) = e(H) - e(G_1[U, B \setminus B']) \stackrel{(6.5)}{\geq} \frac{d|G_1|}{40} - \frac{d(|G_1| - |B'|)}{100} \geq \frac{d|G_1|}{100} = \frac{dt|U|}{100}.$$

Since every subgraph of H has average degree at most $3d$, we have

$$\frac{dt|U|}{50(|U| + |B'|)} \leq \frac{2e(G_1[U, B'])}{|U| + |B'|} \leq d(H[U, B']) \leq 3d.$$

Thus by the fact that $t = L/10$ and $1/L \ll 1$,

$$|B'| \geq \frac{(t-150)|U|}{150} \geq \frac{t|U|}{200}. \tag{6.6}$$

Let $B'' \subseteq B'$ be a subset of B' with

$$|B''| = t^{1/2}|U| = \frac{|G_1|}{t^{1/2}} \geq \frac{K^2d}{(L/10)^{1/2}} > Kd,$$

where we used Claim 6.4(iv) and that $1/K \ll 1/L$. Now, for each such B'' , we define a cycle $C_{B''}$. Let $H' := H[B'', U]$. Then again, since $d_H(v) \geq d/100$ for all $v \in B''$,

$$d(H') = \frac{2e(H')}{|H'|} \geq \frac{d|B''|/100}{|B''| + |U|} \geq \frac{d|B''|/100}{2|B''|} \geq \frac{d}{200}.$$

Thus by Theorem 3.3, H' contains a cycle of length at least $d/200$. Let $C_{B''}$ be any such cycle, then we have $(V(C_{B''}) \setminus U) \subseteq B''$ and $|V(C_{B''}) \cap B'| \geq d/400$.

For a cycle C in H , $V(C) = V(C_{B''})$ holds for at most

$$\binom{|B'| - |V(C) \cap B'|}{|B''| - |V(C) \cap B'|} \leq \binom{|B'| - d/400}{t^{1/2}|U| - d/400}$$

distinct subsets $B'' \subseteq B'$ of size $t^{1/2}|U|$. Thus, we get

$$\begin{aligned} c(G) \geq c(H) &\geq \sum_{|B''|=t^{1/2}|U|} \binom{|B'| - d/400}{t^{1/2}|U| - d/400}^{-1} = \binom{|B'|}{t^{1/2}|U|} \binom{|B'| - d/400}{t^{1/2}|U| - d/400}^{-1} \\ &= \frac{|B'|(|B'| - 1) \dots (|B'| - d/400 + 1)}{t^{1/2}|U|(t^{1/2}|U| - 1) \dots (t^{1/2}|U| - d/400 + 1)} \stackrel{(6.6)}{\geq} \left(\frac{t^{1/2}}{400}\right)^{\frac{d}{400}} \geq \left(\frac{L^{1/2}}{4000}\right)^{\frac{d}{400}} \geq 2^{2d}, \end{aligned}$$

a contradiction. Note that we get the final inequality since $1/L \ll 1$. Thus $d(G_1[B]) \geq 0.9(d - 1)$, as required. \blacksquare

Define

$$U := \{v \in G_1 : d_{G_1}(v) \geq Ld\}.$$

Then

$$|U| \leq 5|G_1|/L, \tag{6.7}$$

otherwise

$$d(G_1) \geq \frac{Ld|U|}{|G_1|} \geq 5d,$$

contradicting Claim 6.4(ii). Clearly, $\Delta(G_1 - U) \leq Ld$, and by Claim 6.5,

$$d_1 := d(G_1 - U) \geq 0.9(d - 1) \geq 0.89d. \tag{6.8}$$

Observe that $\varepsilon_1 \leq \frac{1}{130}$ and

$$0.89 \left(1 - \frac{13\varepsilon_1}{\log 3}\right) \stackrel{(6.1)}{>} 0.8 \quad \text{and} \quad \frac{\varepsilon_1 d_1}{6 \log^2(5/(\frac{1}{30}))} \stackrel{(6.8)}{\geq} \frac{0.89\varepsilon_1 d_0}{6 \log^2(150)} \stackrel{(6.1)}{\geq} 2.$$

Thus we can apply Lemma 5.2 to $G_1 - U$ with $13, \varepsilon_1, 1/30$ and d_1 playing the roles of C, ε_1, c' and d respectively to find a 2-connected $(\varepsilon_1, d_1/30)$ -expander $H_1 \subseteq G_1 - U$ such that

$$d(H_1) \geq 0.8d; \quad \delta(H_1) \geq 0.4d; \quad \text{and} \quad \Delta(H_1) \leq Ld.$$

Since $1/d, 1/K \ll \varepsilon_1, 1/L \leq 1$ from (6.1), if $|H_1| \geq Kd$, we can apply Lemma 5.4 with d, ε_1 and L playing the roles of d, ε_1 and L respectively to see that

$$c(G) \geq c(H_1) \geq 2^{50d}, \tag{6.9}$$

a contradiction. Thus we have $|H_1| \leq Kd$ and consequently (6.1) and Claim 6.4(iv) imply that

$$|U \cup V(H_1)| \stackrel{(6.7)}{\leq} \frac{5|G_1|}{L} + Kd \leq \left(\frac{5}{L} + \frac{1}{K}\right)|G_1| \leq \frac{10|G_1|}{L}.$$

Thus Claim 6.5 applied with $U \cup V(H_1)$ playing the role of U implies that $d_2 := d(G_1 - U - H_1) \geq 0.9(d - 1) \geq 0.89d$. Apply Lemma 5.2 to $G_1 - U - H_1$ with $13, \varepsilon_1, 1/30$ and d_2 playing the roles of C, ε_1, c' and d respectively to see, as before, that $G_1 - U - H_1$ contains a 2-connected $(\varepsilon_1, d_2/30)$ -expander H_2 such that

$$d(H_2) \geq 0.8d; \quad \delta(H_2) \geq 0.4d; \quad \text{and} \quad \Delta(H_2) \leq Ld.$$

In a similar way as we obtained (6.9), we may assume $|H_2| \leq Kd$.

We have shown that G_1 contains two vertex-disjoint expanders H_1, H_2 . Recall that G_1 is 2-connected by Claim 6.4(i). Thus we can choose two minimal vertex-disjoint paths P_1 and P_2 from $V(H_1)$ to $V(H_2)$ in G_1 . Assume that P_1 is from $x_1 \in V(H_1)$ to $x_2 \in V(H_2)$ and P_2 is from $y_1 \in V(H_1)$ to $y_2 \in V(H_2)$.

Then for any path P in H_1 from x_1 to y_1 and any path P' in H_2 from x_2 to y_2 , we obtain a cycle $C(P, P')$ with edge-set $E(P) \cup E(P') \cup E(P_1) \cup E(P_2)$. That is,

$$V(P) \cup V(P') \cup V(P_1) \cup V(P_2)$$

forms a Hamiltonian subset. Thus $c(G) \geq p_{x_1 y_1}(H_1) \cdot p_{x_2 y_2}(H_2)$. Applying Lemmas 4.1 and 4.2 to H_i , for $i \in [2]$, and with $0.8d$ playing the role of d , we get

$$c(G) \geq p_{x_1 y_1}(H_1) \cdot p_{x_2 y_2}(H_2) \geq 2^{0.89 \cdot 0.8d} \cdot 2^{0.89 \cdot 0.8d} \geq 2^{1.4d},$$

a contradiction. Thus the counterexample does not exist for any $d \geq d_0$ and we have shown that, in this range, $d(G) \geq (2 - \alpha)2^{d+1}$ as long as $d(G) \geq d$ and $G \notin \{K_{d+1}, K_{d+1} * K_d\}$. This completes the proof of Theorem 1.3. \square

7. CONCLUDING REMARKS

In this paper, we proved Conjecture 1.1 for all large d . The most obvious remaining open question is to extend this to all d . As we rely on the regularity lemma, we cannot hope to apply our techniques here.

Our proof of Komlós's conjecture can also be adapted to prove Theorem 1.4, the bipartite version of Conjecture 1.1 asked by Tuza [35]. We only sketch the proof here since it is very similar to the proof of Theorem 1.3. Indeed, proceed exactly as in the proof of Theorem 1.3 in Section 6, with Lemmas 4.1 and 4.2 replaced by Lemma 4.1' and 4.2' stated below. (In fact much of this, though true, is redundant here since we are now assuming that $\delta(G) \geq d$.)

Let $a \leq b$ be positive integers. Every cycle in $K_{a,b}$ contains exactly i vertices from each class for some $2 \leq i \leq a$. Thus

$$c(K_{a,b}) = \sum_{i=2}^a \binom{a}{i} \binom{b}{i} = \sum_{i=0}^a \binom{a}{a-i} \binom{b}{i} - \binom{a}{a} \cdot 1 - \binom{a}{a-1} \binom{b}{1} = \binom{a+b}{a} - (ab + 1)$$

and in particular

$$c(K_{d,d}) = \binom{2d}{d} - (d^2 + 1) \sim \frac{2^{2d}}{\sqrt{\pi d}}.$$

We can use Lemma 5.4 in its present form since it guarantees 2^{50d} Hamiltonian subsets, which is still sufficient for Theorem 1.4.

Lemma 4.1'. *Let $\alpha > 0$. Then there exists $d_0 > 0$ such that the following holds for all $d \geq d_0$. Let $n \in \mathbb{N}$ be such that $2d \leq n \leq 2.2d$. Let G be an n -vertex bipartite graph with $\delta(G) \geq d$. Then*

- (i) $c(G) \geq (1 - \alpha) \binom{n}{d} \geq (1 - \alpha) c(K_{d,n-d})$;
- (ii) if $x, y \in V(G)$ are distinct, then $p_{xy}(G) \geq (1 - \alpha) \binom{n-2}{d-1}$.

Observe that $c(K_{d,n-d})$ is at least almost twice as large as $c(K_{d,d})$ when $n > 2d$, and $p_{x'y'}(K_{d,n-d}) \geq \binom{n-2}{d-1}$ for any distinct vertices $x', y' \in V(K_{d,n-d})$.

Lemma 4.2'. *For all $K > 0$, there exists $d_0 = d_0(K)$ such that the following holds for all $d \geq d_0$. Let G be a 2-connected bipartite graph on n vertices where $2.19d \leq n \leq Kd$ and $\delta_2(G) \geq d$. Let x, y be two distinct vertices of G . Then*

$$c(G) > 2^{(2 + \frac{1}{200})d} \quad \text{and} \quad p_{xy}(G) > 2^{1.89d}. \quad (7.1)$$

The proof of Lemma 4.1' is very similar to its original counterpart. Any graph satisfying the hypotheses is an almost balanced almost complete bipartite graph and thus a similar probabilistic argument shows that almost every balanced subset is Hamiltonian. (A version of Dirac's theorem states that any balanced bipartite graph with parts of order k and minimum degree at least $(k+1)/2$ is Hamiltonian.) To prove Lemma 4.2', one must simply replace Lemma 4.3 with the following statement.

Lemma 4.3'. *Let $n \geq 2.18d$ and let G be an n -vertex bipartite graph with $\delta(G) \geq d$. Then G contains at least one of the following:*

- (i) two vertex-disjoint cycles C_1, C_2 with $|C_1| + |C_2| \geq 3.8d$;
- (ii) a path P with $|P| \geq (2 + 1/100)d$;
- (iii) an (a, b) -sun with $a \geq 2d$ and $b \geq d/20$.

Again, the proof is very similar.

In a graph G , say that $U \subseteq V(G)$ is a *weak Hamiltonian subset* of G if $G[U]$ contains a spanning cycle; or $U = \{x, y\}$ and $xy \in E(G)$; or $|U| \in \{0, 1\}$. Observe that every subset of the vertices of a complete graph is weak Hamiltonian. Theorem 1.3 immediately implies that there exists $d_0 > 0$ such that for all integers $d \geq d_0$, every graph G with average degree at least d contains at least as many weak Hamiltonian subsets as K_{d+1} (that is, 2^{d+1}). This proves another conjecture of Tuza [35], stated in the quest for ‘nicer formulas’ than (1.1).

8. ACKNOWLEDGEMENTS

We are indebted to the referee for her/his careful reading and valuable comments.

REFERENCES

- [1] W. Ahrens, Über das Gleichungssystem einer Kirchhoffschen galvanischen Stromverzweigung, *Math. Ann.* **49** (1897), 311–324.
- [2] A. Arman, D. S. Gunderson and S. Tsaturian, Triangle-free graphs with the maximum number of cycles, *Discrete Math.* **339** (2) (2016), 699–711.
- [3] R. E. L. Aldred and C. Thomassen, On the number of cycles in 3-connected cubic graphs, *J. Combin. Theory B* **71** (1997), 79–84.
- [4] J. Balogh, H. Liu and M. Sharifzadeh, Subdivisions of a large clique in C_6 -free graphs. *J. Combin. Theory Ser. B*, **112**, (2015), 18–35.
- [5] C. A. Barefoot, L. Clarke and R. Entringer, Cubic graphs with the minimum number of cycles, *Congr. Numer.* **53** (1986), 49–62.
- [6] B. Csaba, On the Bollobás-Eldridge conjecture for bipartite graphs, *Combin. Probab. Comput.* **16** (2007), 661–691.
- [7] B. Csaba, D. Kühn, A. Lo, D. Osthus, A. Treglown, Proof of the 1-factorization and Hamilton decomposition conjectures, *Mem. Amer. Math. Soc.*, to appear.
- [8] B. Cuckler and J. Kahn, Hamiltonian cycles in Dirac graphs, *Combinatorica* **29** (2009), 299–326.
- [9] G. A. Dirac, Some theorems on abstract graphs, *Proc. London Math. Soc.* **2** (1952), 69–81.
- [10] P. Erdős and T. Gallai, On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hunar.* **10** (1959), 337–356.
- [11] S. Janson, T. Łuczak and A. Ruciński, *Random graphs*, Wiley-Interscience, 2000.
- [12] R. Karp, Reducibility among combinatorial problems, *Complexity of computer computations*, Plenum, New York, (1972), 85–103.
- [13] J. Kim, D. Kühn, D. Osthus, M. Tyomkyn, *A blow-up lemma for approximate decompositions*, submitted, arXiv:1604.07282 (2016).
- [14] M. Krivelevich and W. Samotij, Optimal packings of Hamilton cycles in sparse random graphs, *SIAM J. Disc. Math.* **26** (3) (2012), 964–982.
- [15] M. Knor, On the number of cycles in k -connected graphs, *Acta Math. Univ. Comen.* **63** (2), (1994), 315–321.
- [16] F. Knox, D. Kühn and D. Osthus, Edge-disjoint Hamilton cycles in random graphs, *Random Struct. Alg.* **46** (2015), 397–445.
- [17] J. Komlós, G.N. Sárközy and E. Szemerédi, Blow-up lemma, *Combinatorica* **17** (1997), 109–123.
- [18] J. Komlós, A. Shokoufandeh, M. Simonovits and E. Szemerédi, The regularity lemma and its applications in graph theory, *Theoretical aspects of computer science* (Tehran, 2000), 84–112, Lecture Notes in Comput. Sci., 2292, Springer, Berlin, 2002.
- [19] J. Komlós and M. Simonovits, Szemerédi’s regularity lemma and its applications in graph theory, *Combinatorics: Paul Erdős is eighty, Vol. 2* (Keszthely, 1993), 295–352, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.
- [20] J. Komlós, E. Szemerédi, Topological cliques in graphs. *Combin. Probab. Comput.*, **3**, (1994), 247–256.
- [21] J. Komlós, E. Szemerédi, Topological cliques in graphs II. *Combin. Probab. Comput.*, **5**, (1996), 79–90.
- [22] D. Kühn, D. Osthus, Hamilton decompositions of regular expanders: a proof of Kelly’s conjecture for large tournaments, *Adv. Math.*, **237**, (2013), 62–146.
- [23] D. Kühn, D. Osthus, Hamilton decompositions of regular expanders: applications, *J. Comb. Theory B* **104** (1) (2014), 1–27.
- [24] D. Kühn, D. Osthus, Hamilton cycles in graphs and hypergraphs: an extremal perspective, *Proceedings of the International Congress of Mathematicians*, (2014), Seoul, Korea, Vol 4, 381–406.
- [25] D. Kühn, D. Osthus and A. Taraz, Large planar subgraphs in dense graphs, *J. Combin. Theory B* **95** (2005), 263–281.
- [26] H. Liu, M. Sharifzadeh and K. Staden, Local conditions for exponentially many subdivisions, *Combin. Probab. Comput.*, to appear.
- [27] H. Liu, R.H. Montgomery, A proof of Mader’s conjecture on large clique subdivisions in C_4 -free graphs. *J. London Math. Soc.*, to appear.

- [28] R.H. Montgomery, Logarithmically small minors and topological minors. *J. Lond. Math. Soc.*, **91** (2), (2015), 71–88.
- [29] N. Morrison and A. Scott, Maximising the number of induced cycles in a graph, submitted, arXiv:1603.02960 (2016).
- [30] W. Perkins, personal communication.
- [31] L. Pósa, A theorem concerning Hamiltonian lines, *Magyar Tud. Akad. Mat. Fiz. Oszt. Kozl.* **7** (1962), 225–226.
- [32] E. Szemerédi, Regular partitions of graphs, *Problèmes Combinatoires et Théorie des Graphes Colloques Internationaux CNRS* **260** (1978), 399–401.
- [33] Zs. Tuza, Exponentially many distinguishable cycles in graphs, *J. Combin. Infor. Sys. Sci.* **15** (1–4), (1990), 281–285.
- [34] Zs. Tuza, Unsolved combinatorial problems, Part I, *BRICS Lecture Series, LS-01-1*, (2001).
- [35] Zs. Tuza, Problems on cycles and colorings, *Disc. Math.* **313** (2013), 2007–2013.
- [36] L. Volkmann, Estimations for the Number of Cycles in a Graph, *Per. Math. Hungar.* **33**, (1996), 153–161.

| | |
|---|--|
| Jaehoon Kim School of Mathematics University of Birmingham Birmingham B15 2TT UK | Hong Liu, Maryam Sharifzadeh and Katherine Staden Mathematics Institute University of Warwick Coventry CV4 7AL UK |
|---|--|

E-mail addresses: j.kim.3@bham.ac.uk,
 {h.liu.9, m.sharifzadeh, k.l.staden}@warwick.ac.uk.

APPENDIX A.

Here we give the details of the omitted proofs of Proposition 4.4 and Lemma 4.9.

Proof of Proposition 4.4. It is convenient to prove a general claim from which (i)–(iii) will follow. Let G be a graph with $V(G) = \{z_1, \dots, z_\ell\}$. Given $k \in \mathbb{N}$, let $(z_1, \dots, z_\ell)^k$ denote the sequence

$$(z_1, \dots, z_\ell, z_1, \dots, z_\ell, \dots, z_1, \dots, z_\ell)$$

of length $k\ell$. Consider the following claim.

Claim A.1. *Let $N \in \mathbb{N}$ and let $R := (y_1, \dots, y_\ell)$ be a circuit in G with $V(G) = \{y_1, \dots, y_\ell\}$. Let $x, x' \in V(G)$ be two not necessarily distinct vertices. Then there is a walk W in G from x to x' and $N \cdot \deg(y, R) \leq \deg(y, W) \leq (N + 1) \cdot \deg(y, R)$ for all $y \in V(G)$.*

Proof: Assume, without loss of generality, that x appears no later than x' in R . Then we can choose $1 \leq s \leq t \leq \ell$ such that $y_s = x$ and $y_t = x'$. Define the concatenation

$$W := \begin{cases} (y_t, \dots, y_\ell)(y_1, \dots, y_\ell)^N(y_1, \dots, y_s) & \text{if } s < t \\ (y_t, \dots, y_\ell)(y_1, \dots, y_\ell)^{N-1}(y_1, \dots, y_s) & \text{if } s = t, \end{cases}$$

where (y_1, \dots, y_s) and (y_t, \dots, y_ℓ) are subwalks of R . It is clear by the definition of R that W is a walk in G . Then clearly we have

$$N \cdot \deg(y, R) \leq \deg(y, W) \leq (N + 1) \cdot \deg(y, R),$$

as required. ■

To prove (i), write $C = x_1 \dots x_a$ and define a circuit

$$R := (x_1, \dots, x_a).$$

Then $\deg(x, R) = 1$ for all $x \in V(C)$ and we are done by applying Claim A.1 with $2n$ playing the role of N .

To prove (ii), write $P = x_1 \dots x_a$ and let

$$R := (x_1, \dots, x_a, x_{a-1}, \dots, x_2)$$

be a circuit of length $2a - 1$. Clearly $\deg(x, R) = 1$ for $x \in \{x_1, x_a\}$ and $\deg(x, R) = 2$ for $x \in V(P) \setminus \{x_1, x_a\}$. We are done by applying Claim A.1 with n playing the role of N .

For (iii), write $V(S) := \{x_1, \dots, x_a\} \cup \{y_{i_1}, \dots, y_{i_b}\}$ where $\text{Cor}(S) = \{x_{i_k}, y_{i_k} : k \in [b]\}$. Define R to be the circuit in S obtained by concatenating the two cycles, one with x_{i_j} 's and the other with y_{i_j} 's for $j \in [b]$: that is,

$$R := (x_{i_1}, x_{i_1+1}, \dots, x_a, x_1, \dots, x_{i_1-1}, y_{i_1}, x_{i_1+1}, \dots, x_{i_b-1}, y_{i_b}, x_{i_b+1}, \dots, x_a, x_1, \dots, x_{i_1-1}).$$

It is easy to see that, for $x \in \text{Cor}(S)$ we have $\deg(x, R) = 1$ and for $x \in V(S) \setminus \text{Cor}(S)$ we have $\deg(x, R) = 2$. We are immediately done by applying Claim A.1 with n playing the role of N . \square

We now turn to the proof of Lemma 4.9. Given a graph R with $V(R) = [r]$ and an integer $\ell \leq r$, we say that a graph R' is the $(\ell, 2)$ -blow-up of R if $V(R') = [r + \ell]$ and

$$E(R') = E(R) \cup \bigcup_{i \in [\ell]} (\{i + r, x\} : x \in N_R(i)\} \cup \{i + r, j + r\} : j \in N_R(i) \cap [\ell]\}.$$

That is, R' is obtained from R by duplicating vertices $[\ell] \subseteq V(R)$.

Lemma A.2. *Suppose $0 < 1/n \ll \varepsilon \ll \gamma, 1/\Delta_R \leq 1$ and $1/n \ll 1/r$ and $0 \leq \ell \leq r$ with $r, n, \ell \in \mathbb{N}$. Let R be a graph with $V(R) = [r]$ and $\Delta(R) \leq \Delta_R$. Let G be an (ε, γ) -super-regular graph with respect to a vertex partition (R, V_1, \dots, V_r) such that*

$$2n \leq |V_i| \leq 2n + 2 \text{ for } i \in [\ell] \quad \text{and} \quad n \leq |V_i| \leq n + 2 \text{ for } \ell + 1 \leq i \leq r.$$

Let R' be the $(\ell, 2)$ -blow-up of R . For $i \in [r] \setminus [\ell]$, let $V'_i := V_i$. Then for all $i \in [\ell]$, there exists a partition $V_i = V'_i \cup V'_{r+i}$ such that $n \leq |V'_i|, |V'_{r+i}| \leq n + 2$ and G is $(4\varepsilon, \gamma/2)$ -super-regular with respect to vertex partition $(R', V'_1, \dots, V'_{r+\ell})$.

The proof of this lemma is a standard application of slicing lemma and Chernoff bounds. We provide only a sketch here: For each $i \in [\ell]$, we take a random partition $V'_i \cup V'_{r+i}$ of V_i such that $|V'_i| = n$. Then slicing lemma ensures that $G[V'_i, V'_j]$ is $(4\varepsilon, \gamma/2)$ -regular for each $ij \in E(R')$. For a vertex $v \in V_i$ and $j \in N_{R'}(i)$, Chernoff bounds give us that

$$\mathbb{P}[|d_G(v) \cap V'_j| \geq \gamma n/2] \geq 1 - e^{-\varepsilon n^2}.$$

Thus, union bounds over all vertices v and all $i, j \in [r + \ell]$ show that the conclusion holds with high probability.

Proof of Lemma 4.9. Let

$$N_1 := \{i \in [r] : n \leq n_i \leq n + 2\} \quad \text{and} \quad N_2 := \{i \in [r] : 2n \leq n_i \leq 2n + 2\}.$$

So $[r] = N_1 \cup N_2$ is a partition. Without loss of generality, we may assume that $N_2 = \{1, \dots, \ell\}$ for some $0 \leq \ell \leq r$. Let R' be the $(\ell, 2)$ -blow-up of R . Apply Lemma A.2 to find a new partition $V'_1, \dots, V'_{r+\ell}$ of G such that G is $(4\varepsilon, \gamma/2)$ -super-regular with respect to vertex partition $(R', V'_1, \dots, V'_{r+\ell})$ such that $|V'_i| \in \{n, n + 1, n + 2\}$ for each $i \in [r + \ell]$. Let $W' := (w'_1, \dots, w'_m)$ be a walk in R' obtained as follows. For all $t \in [m]$, let $w'_t := w_t$ if $w_t \in N_1$, and for each w'_t such that $w_t \in N_2$, we choose $w'_t \in \{w_t, w_t + r\}$ so that $\deg(i, W') = |V'_i|$ and $\deg(i + r, W') = |V'_{i+r}|$ for all $i \in [\ell]$. This is possible by (iv) and fact that $n_i = |V'_i| + |V'_{i+r}|$. Thus $\deg(i, W') = |V'_i|$ for all $i \in [r + \ell]$.

Let $Q = (q_1, \dots, q_m)$ be a path of length m . For each $i \in [r + \ell]$, let $X_i := \{q_j : w'_j = i\}$. Since W' is a walk in R' with $\deg(i, W') = |V'_i| = |X_i|$, Q admits a vertex partition $(R', X_1, \dots, X_{r+\ell})$. Let $G' := G - \{x, y\}$. Then G' is $(5\varepsilon, \gamma/3)$ -super-regular with respect to vertex partition $(R', V''_1, \dots, V''_{r+\ell})$, where $V''_i := V'_i \setminus \{x, y\}$ for all $i \in [r + \ell]$. Let $S_{q_2} := N_{G'}(x, V''_{w'_2})$ and $S_{q_{m-1}} := N_{G'}(y, V''_{w'_{m-1}})$. Now $w'_1 w'_2 \in E(R')$ by definition. Since $G'[V''_{w'_1}, V''_{w'_2}]$ is $(5\varepsilon, \gamma/3)$ -super-regular, we have that $|S_{q_2}| \geq \gamma |V''_{w'_2}|/3$. Similarly $|S_{q_{m-1}}| \geq \gamma |V''_{w'_{m-1}}|/3$.

Let Q^* denote the truncated path (q_2, \dots, q_{m-1}) . Theorem 4.8 implies that there is an embedding of Q^* into G' such that q_2 is mapped to a vertex in S_{q_2} and q_{m-1} is mapped to a vertex in $S_{q_{m-1}}$ and $V(Q^*) = V(G')$. By the choice of S_{q_2} and $S_{q_{m-1}}$, this embedding can be extended to x and y to obtain a path P which spans $V(G)$ and has endpoints x and y . \square